

O'Bryen, Barbara

From: Swope, Sheridan
Sent: Wednesday, December 31, 2003 1:40 PM
To: O'Bryen, Barbara
Subject: FW: 09980,818

Importance: High 09-980881

[Swope, Sheridan] -----Original Message-----

From: Swope, Sheridan
Sent: Monday, December 22, 2003 5:02 PM
To: STIC-Biotech/ChemLib
Subject: 09980,818

For 09980,818, pls search and interference search:

SID 1: Full-length and nt18-1097 against the NT and AA data bases
SID 2: against the NT and AA data bases
SID 3: against the NT and AA data bases
SID 4: against the NT and AA data bases
SID 9: against the NT and AA data bases.

THANKS!!!

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
sheridan.swope@uspto.gov
703-305-1696 (voice)
703-308-3014 (FAX)
Mailbox: CM1 Rm10D01
Office: CM1 Rm12D12

As of January 12, 2003:
571-272-0943 (voice & FAX)
E03A70 Remsen Bld (Office)
E03A81 Remsen Bld (Mailbox)
400 Dulany Street
Alexandria, VA

Scan this Page

*Barb O'Bryen
01/08/04*

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:38:01 ; Search time 206.037 Seconds
(without alignments)
2423.606 Million cell updates/sec

Title: US-09-980-881a-1
Perfect score: 2777
Sequence: 1 agaaattgctgtggtggtg.....aaaaa.....1573

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135747_24935/app_query.fasta_1.2958
-DB-A_Geneseq_19Jun03 -QPMF=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 @CGN 1.117 @runat_31122003_135747_24935 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1911	68.8	360	21	AA811457 Human brain carbox
2	1907.5	68.7	386	24	ABG72095 Human protease pro
3	1902.5	68.5	386	24	ABG72093 Human protease pro
4	1885	67.9	423	17	AA90293 Human plasma carbo
5	1885	67.9	423	18	AAW14733 Human plasma carbo
6	1885	67.9	423	20	AAW22270 Human plasma carbo
7	1880	67.7	423	14	AA836273 Human plasma carbo
8	1785	64.3	338	21	AA811458 Human brain carbox
9	1325	47.7	246	21	AA811459 Human brain carbox
10	889	32.0	211	21	AA858129 Lung cancer associ
11	652.5	23.5	417	24	AB996888 Amino acid sequenc
12	651.5	23.5	415	17	AAW06172 Human pancreatic c
13	651.5	23.5	415	19	AAW74476 Human pancreatic c
14	647.5	23.3	417	23	AAU84325 Protein CPA3 diffe
15	647.5	23.3	667	24	ABP97414 Anti-human seminal
16	644	23.2	424	17	AAW06175 ProHCPB with PelB
17	640.5	23.1	404	16	AAW75131 Porcine Tyr-His-Me
18	636	22.9	324	22	AA66562 Bothops jarasaca c
19	635.5	22.9	437	22	AAE00502 Human carboxypepti
20	635.5	22.9	437	23	AAE19175 Human protease, PR
21	634.5	22.8	716	18	AAW41414 PreproHCPB-linker-
22	630.5	22.7	434	24	AB996887 Amino acid sequenc
23	607.5	21.9	307	17	AAW0602 Rat carboxypeptida
24	576	20.7	306	16	AAW75132 Porcine carboxypep
25	576	20.7	306	22	AAU04477 Porcine carboxypep
26	575	20.7	349	17	AAW06173 Mature HCPB with p
27	572.5	20.6	329	17	AAW41412 Humanised Fd-mutan
28	571.5	20.6	613	19	AAW74481 Human pancreatic c
29	571.5	20.6	613	19	AAW74481 Human pancreatic c
30	570	20.5	354	22	AA66566 Human secreted met
31	570	20.5	362	22	AA66566 Human secreted met
32	570	20.5	374	22	AA66547 Human protease and
33	570	20.5	374	22	AA66547 Human secreted met
34	570	20.5	374	22	AA66565 Amino acid sequenc
35	570	20.5	374	23	AAU82703 NOX protein sequen
36	570	20.5	374	24	ABJ37888 Human zinc-binding
37	570	20.5	374	24	ABG74658 Modified HCPB (D25
38	568	20.5	349	17	AAW06181 Modified HCPB (D25
39	568	20.5	349	17	AAW13749 Modified HCPB (D25
40	567	20.4	349	18	AAW06182 Carboxypeptidase B
41	567	20.4	349	18	AAW13750 Carboxypeptidase B
42	564	20.3	349	18	AAW13763 Carboxypeptidase B
43	563	20.3	444	23	AAE24340 Human 23566 (carbo
44	562	20.2	349	18	AAW13752 Carboxypeptidase B
45	562	20.2	349	18	AAW13757 Carboxypeptidase B

ALIGNMENTS

RESULT 1

AA811457

ID AA811457 standard; Protein; 360 AA.

AC AA811457;

XX 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B protein.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; antialzheimers; nootropic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma.

XX Homo sapiens.

XX

XX (GANW/) GAN W.
 PA (KETC/) KETCHUM K A.
 PA (DPRA/) DI FRANCESCO V.
 PA (BEAS/) BEASLEY E M.
 XX
 PI Gan W, Ketchum KA, Di Francesco V, Beasley EM;
 XX
 DR WPI; 2003-102387/09.
 DR N-PSDB; ABX13670, ABX13671.
 XX
 PT New isolated human protease proteins, useful for developing therapeutic
 PT or diagnostic compositions, particularly for developing human
 PT therapeutic agents that modulate protease activity in cells or tissues
 PT
 XX
 PS Claim 1; Page -: 80pp; English.
 CC
 CC The invention discloses an isolated human protease peptide, its allelic
 CC variant or orthologue. The proper functioning of the cell requires
 CC careful control of the levels of important structural proteins, enzymes
 CC and regulatory proteins. One of the ways the cell reduces the steady
 CC state level of a particular protein is by proteolytic degradation.
 CC Proteolysis can also be used to convert a pre or pro-protein in to an
 CC active form. Proteases also regulate many different cell proliferation,
 CC differentiation and signalling processes. The peptides and nucleic acid
 CC molecules are useful in the development of human therapeutics (gene and
 CC protein therapy) and diagnostic compositions. The peptides are also
 CC useful for raising antibodies or eliciting an immune response (vaccine),
 CC as a reagent (including the labeled reagent) in assays designed to
 CC quantitatively determine levels of the protein (or its binding partner or
 CC ligand) in biological fluids, or as markers for tissues in which the
 CC corresponding protein is preferentially expressed and in methods for
 CC identifying a modulator of the peptide or an agent that binds to the
 CC peptide. The agents identified are useful for treating protease-related
 CC conditions that are specific for the subfamily of proteases that the
 CC peptide belongs to, particularly in cells and tissues that express the
 CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
 CC disorders. The modulator of the peptide is also useful for treating a
 CC disorder characterised by an absence of, inappropriate or unwanted
 CC expression of the protein. The sequence presented is the human protease
 CC protein, variant #2, the gene for which is located on chromosome 13.
 CC Note: This sequence is not shown in the specification but has been
 CC created by the indexer from information given in figure 3.
 XX
 SQ Sequence 386 AA;
 Alignment Scores:
 Pred. No.: 1.15e-196 Length: 386
 Score: 1907.50 Matches: 364
 Percent Similarity: 94.56% Conservative: 1
 Best Local Similarity: 94.30% Mismatches: 4
 Query Match: 68.69% Indels: 18
 DB: 24 Gaps: 1
 US-09-980-881A-1 (1-1573) x ABG72095 (1-386)
 QY 18 ATGAAGCTTTCAGAGCTTGAGTCCAGTCTGTACCATCTTCTCTGTGAGCAGCATGTC 77
 DB 1 MetCysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
 QY 78 TTCGCGTTCAGAGTGCAGGTTCTAGCTCTTCTTCTAGAACCTCTAGGCAAGTTCAA 137
 DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
 QY 138 GTTCTACAGATCTTACTACAAATAGCATGAGATTGTTCTCTGGCAGCGCGTAAACAGCTGAC 197
 DB 41 ValLeuGlnAsnLeuThrThrThrGluValLeuValLeuValLeuValLeuValLeuVal 60
 QY 198 CTTATTGTGAAGAAAACAGTCCATTTTGTAAATGCATCTCATGTCCGACATGTG 257
 DB 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

QY 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGACAGTGTCTTCTGCGACAGCTGGAGAT 317
 DB 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
 QY 318 CTTATTCAACAGCAGATTTTCAACACAGACAGTCCAGCCCGAGCTCCCGCATCGTACTAT 377
 DB 101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
 QY 378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTAGAGGCAT 437
 DB 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
 QY 438 CTTGATATCTTACAAAAATCCACATGATCTCTATTTGAGAGTACCCACTCTATGTT 497
 DB 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
 QY 498 TTAAGGTTTTCGAAAAAGAACAAACAGCAAAATGCCATATGGATTGACTCTCGAATC 557
 DB 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
 QY 558 CATGCCAGAGATGGATCTCTCTGCTTCTGCTGTGTTTCATAGGCCCATTAATCGAATG 617
 DB 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
 QY 618 TGGAGAAAGAACCGTTCTTCTATGCGAACATCATTCGATCGGAACAGACCTGAATAGC 677
 DB 201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220
 QY 678 AACTTTGTCTCAAAACACATGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACC 737
 DB 221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240
 QY 738 TACTGTGGACTTTTATCTGAGTCAGAACCCAGAGTGAAGCGAGTGGCTGTCTTCTTGAGA 797
 DB 241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
 QY 798 AGAATATCAACAGATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTG 857
 DB 261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
 QY 858 TTTCCATATTTCTATACAGTAGTAAACAGACCATGAGCACTGTCTCTAGTAGCC 917
 DB 281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
 QY 918 AGTGAAGCAGTTCGTGCTATTGACAAAATAGTAAATAATACAGGTATATACATGCGCAT 977
 DB 301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320
 QY 978 GGCTCAGAAACCTTATACCTAGCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGGC 1037
 DB 321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340
 QY 1038 ATCAATATTCGTT----- 1051
 DB 341 IleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360
 QY 1052 -----TACATCAAAACCCACCTGTAGAGAGCTTTTTCGCGCTGTCTCTCTAAAATAGTTGG 1105
 DB 361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaValSerLysIleAlaTrp 380
 QY 1106 CATGTCTAGGAATGTT 1123
 DB 381 HisValIleArgAsnVal 386
 RESULT 3
 ABG72093
 ID ABG72093 standard; Protein; 386 AA.
 XX
 AC ABG72093;
 XX
 DT 14-FEB-2003 (first entry)
 XX
 DE Human protease protein, variation #1.

XX Human; enzyme; protease; proteolytic degradation; proteolysis;
 KW proliferation; differentiation; signalling; therapeutic; gene therapy;
 KW protein therapy; diagnostic; immune response; vaccine; inflammation;
 KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
 KW single nucleotide polymorphism; SNP.
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 169
 FT /note= "This residue is the major variant of a single
 FT nucleotide polymorphism. The minor variant is an
 FT Thr residue"
 XX
 PN US2002137179-A1.
 XX
 XX 26-SEP-2002.
 XX
 PF 21-MAR-2001; 2001US-0813133.
 XX
 PR 21-MAR-2001; 2001US-0813133.
 XX
 PA (GANW/) GAN W.
 PA (KETC/) KETCHUM K A.
 PA (DFRA/) DI FRANCESCO V.
 PA (BEAS/) BEASLEY E M.
 XX
 PI Gan W, Ketchum KA, Di Francesco V, Beasley EM;
 XX WPI; 2003-102387/09.
 DR N-PSDB; ABX13670, ABX13671.
 DR
 XX New isolated human protease proteins, useful for developing therapeutic
 FT or diagnostic compositions, particularly for developing human
 FT therapeutic agents that modulate protease activity in cells or tissues
 FT
 PT
 PT
 XX
 XX Claim 1; Fig 2; 80pp; English.
 XX
 CC The invention discloses an isolated human protease peptide, its allelic
 CC variant or orthologue. The proper functioning of the cell requires
 CC careful control of the levels of important structural proteins, enzymes
 CC and regulatory proteins. One of the ways the cell reduces the steady
 CC state level of a particular protein is by proteolytic degradation.
 CC Proteolysis can also be used to convert a pre or pro-protein in to an
 CC active form. Proteases also regulate many different cell proliferation,
 CC differentiation and signalling processes. The peptides and nucleic acid
 CC molecules are useful in the development of human therapeutics (gene and
 CC protein therapy) and diagnostic compositions. The peptides are also
 CC useful for raising antibodies or eliciting an immune response (vaccine),
 CC as a reagent (including the labeled reagent) in assays designed to
 CC quantitatively determine levels of the protein (or its binding partner or
 CC ligand) in biological fluids, or as markers for tissues in which the
 CC corresponding protein is preferentially expressed and in methods for
 CC identifying a modulator of the peptide or an agent that binds to the
 CC peptide. The agents identified are useful for treating protease-related
 CC conditions that are specific for the subfamily of proteases that the
 CC peptide belongs to, particularly in cells and tissues that express the
 CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
 CC disorders. The modulator of the peptide is also useful for treating a
 CC disorder characterised by an absence of, inappropriate or unwanted
 CC expression of the protein. The sequence presented is the human protease
 CC protein, variant #1, the gene for which is located on chromosome 13.
 XX
 SQ Sequence 386 AA;
 XX
 Alignment Scores:
 Pred. No.: 3,99e-196 Length: 386
 Score: 1902.50 Matches: 363
 Percent Similarity: 94.30% Conservative: 1
 Best Local Similarity: 94.04% Mismatches: 5
 Query Match: 68.51% Indels: 18

DB: 24 Caps: 1
 US-09-980-881A-1 (1-1573) x ABG72093 (1-386)
 QY 18 ATGAAGCTTTGACGCTTGCAGCCTTGCAGTCCCTTGTACCCATTGTTCTCTTGTGTGAGCAGCATGTC 77
 DB 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
 QY 78 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGCTCTCTTCTAGAACCTCTAGGCGAAGTCAA 137
 DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
 QY 138 GTTCTACAGAACTTACTACTACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
 DB 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
 QY 198 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAAATGCATCTCTGATGTCGAATGTG 257
 DB 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
 QY 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGTCAGTGTCTTGTGTCGACACGCTGGAGAT 317
 DB 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
 QY 318 CTTATTCAACAGCAGATTTTCAACAGACACAGTCAGCCCGAGCCCTCCGATCGTACTAT 377
 DB 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
 QY 378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTTAACTGAGAGCAT 437
 DB 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
 QY 438 CTTGATATGCTTACAAAAATCCACATTTGGATCTCTTGTGAGAGTACCCACTCTATGT 497
 DB 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
 QY 498 TTAAGGTTTCTGGAAGAAGCAACAAACAGCAAAATGCCATATGGATTGACTCTGGAATC 557
 DB 161 LeuLysValSerGlyLysGluGlnAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
 QY 558 CATGCCAGAGAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGGCCATATTCGAATG 617
 DB 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
 QY 618 TGGAGAAAGAACCGTTCTTCTATGCGAACATCATGTGATCGGAACACACCTGAATAGC 677
 DB 201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220
 QY 678 AACTTTGTCTCAAAACACCTGGTGTGAGGAAGTGCATCCAGTTCTCTCATGCTCGGAAC 737
 DB 221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240
 QY 738 TACTGTGGACTTTATCTCTGAGTCAGAACAGAGTGAAGCGAGTGCTAGTTCTTTGAGA 797
 DB 241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
 QY 798 AGAATATCAACAGATTAAAGCATACATCAGCATCATTCATCTACTCCAGCATATAGTG 857
 DB 261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
 QY 858 TTTCCATATTCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCC 917
 DB 281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuAla 300
 QY 918 AGTGAAGCAGTTCGCTGCTATTTCACAAAACCTAGTAAATAATACCAGGTATACATGCGCAT 977
 DB 301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320
 QY 978 GGCTCAGAAACCTTATACCTAGCTCTGAGGTGGGAGCATTTGATCTATGATTTGGGC 1037
 DB 321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGly 340
 QY 1038 ATCAATAATTCGTT----- 1051

```

Db      341  ILeIySTySerPheThrIleGluLeuArgAspThrGlyThrTyrglyPheLeuLeuPro 360
QY      1052  -----TACATCAAAACCCTCTAGAGAGCTTTTGCCTGCTCTCTAAATAGCTTGG 1105
Db      361  GluArgTyriLeysProThrCysArgGluAlaPheAlaValSerLysIleAlaTyr 380
QY      1106  CATGTCATTAGGAATGTT 1123
Db      381  HisValIleArgAsnVal 386

RESULT 4
AAR90293
ID      AAR90293 standard; Protein; 423 AA.
XX
AC      AAR90293;
XX
DT      25-MAR-2003 (updated)
DT      12-APR-1996 (first entry)
XX
XX      Human plasma carboxypeptidase B.
DE      Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW      purification; plasminogen; affinity column.
XX
OS      Homo sapiens.
XX
FH      Key
FH      Active-site
FT      181
FT      /note= "Residue suspected of being involved in
FT      catalytic activity of enzyme."
FT
FT      Binding-site
FT      183
FT      /note= "Residue suspected of being part of the
FT      binding site of enzyme."
FT
FT      Binding-site
FT      256..257
FT      /note= "Residues suspected of being part of the
FT      binding site of enzyme."
FT
FT      Active-site
FT      310
FT      /note= "Residue suspected of being involved in
FT      catalytic activity of enzyme."
FT
FT      Binding-site
FT      312
FT      /note= "Residue suspected of being part of the
FT      binding site of enzyme."
FT
FT      Active-site
FT      369
FT      /note= "Residue suspected of being involved in
FT      catalytic activity of enzyme."
FT
FT      Binding-site
FT      370..371
FT      /note= "Residues suspected of being part of the
FT      binding site of enzyme."
FT
FT      Binding-site
FT      394
FT      /note= "Residue suspected of being part of the
FT      binding site of enzyme."
PN
XX      US5474901-A.
XX
PD      12-DEC-1995.
XX
XX      19-JUL-1994; 94US-0277540.
XX
XX      01-FEB-1991; 91US-0649591.
XX      14-OCT-1992; 92US-0959944.
XX      15-DEC-1993; 93US-0167727.
XX      19-JUL-1994; 94US-0277540.
XX
XX      (GETH ) GENENTECH INC.
XX
XX      Drayna DT, Eaton DL;
XX
XX      WPI; 1996-039508/04.
XX      N-PSDB; AAT11671.
XX
XX      Antibody to human plasma carboxypeptidase B - useful for detecting
XX      and purifying hPCPB for use in treating clotting disorders e.g.

```

```

PT      haemophilia A
XX
PS      Disclosure; Figure 4; 40pp; English.
XX
CC      An antibody which specifically binds human plasma carboxypeptidase B
CC      (hPCPB) and does not cross react with other carboxypeptidases is
CC      useful for the detection of hPCPB in vitro. The antibody is also
CC      used for purifying hPCPB from a sample. Purification comprises
CC      passing a sample thought to contain hPCPB over either a column to
CC      which antibody has been bound, or a plasminogen affinity column,
CC      eluting the column and then recovering the fraction containing the
CC      hPCPB.
CC      (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ      Sequence 423 AA;

Alignment Scores:
Pred. No.: 3-24e-194 Length: 423
Score: 1885.00 Matches: 365
Percent Similarity: 86.52% Conservative: 1
Best Local Similarity: 86.29% Mismatches: 3
Query Match: 67.88% Indels: 55
DB: 17 Gaps: 2

US-09-980-881A-1 (1-1573) x AAR90293 (1-423)

QY      18  ATGAAGCTTTGCAGCCTTCAGTCTGTACCCATTGTTCTCTCTGTGAGCAGCATGTC 77
Db      1  MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY      78  TTCGCGTTTCAGAGTGGCCCAAGTTCTAGCTCTCTCTCTAGAACCTCTAGGCGAAGTTCAA 137
Db      21  PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY      138  GTTCTACAGACTTCTTACTACATAGATGAGATTGTTCTCTGCGAGCCGGTAACAGCTGAC 197
Db      41  ValLeuGlnAsnLeuThrThrTyrgluIleValLeuTrpGlnProValThrAlaAsp 60
QY      198  CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGATCTCTGATGTGCAATGTG 257
Db      61  LeuIleValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 80
QY      258  AAGGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTGTGTCGACAGCTGGAAGAT 317
Db      81  LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY      318  CTTATTCAACAGCAGATTTCACACGACAGTCCAGCCCGCCGCTCCGATCGTACTAT 377
Db      101  LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrr 120
QY      378  GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATACTAGAGGCAT 437
Db      121  GluGlnTyrrHisSerLeuAsnGluIleTyrrSerTrpIleGluPheIleThrGluArgHis 140
QY      438  CCTGATATGCTTACAAAATCCACATTCATTCATTTGAGAAAGTACCACCTCTATGTT 497
Db      141  ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrrProLeuTyrrVal 160
QY      498  TTAAGGTTTCTGGAAGAACAACACAGCCCAAAATGCCATATGGATTGACTGTGGAATC 557
Db      161  LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY      558  CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTGTTGTTGTTATAGGCAT----- 608
Db      181  HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY      608  ----- 608
Db      201  PheTyrrGlyIleIleGlyGlnTyrrThrAsnLeuLeuArgLeuValAspPheTyrrValMet 220
QY      609  -----NATCGAATGCGAAG 626
Db      221  ProValValAsnValAspGlyTyrrAspTyrrSerTrpLysLysAsnArgMetTrpArgLys 240

```

QY 627 AACGTTCTTTCTATGCGAACAATCATTGTCATCGAGACACCTGAATAGCAACTTTGTC 686
 Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
 QY 687 TCCAAACACCTGGTGTGAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 746
 Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
 QY 747 CTTTATCTAGTCAGACAGACAGAGTGAAGCGAGTGTGCTAGTTCTTGAGAAATATC 806
 Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
 QY 807 AACAGATTAAAGCATACATCAGCATCTTCTACTCCAGCATATAGTGTTCATAT 866
 Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
 QY 867 TCCTATACGAAGTAAACCAAGACCATGAGAACTGTCTCTAGTAGCAGTGAAGCA 926
 Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
 QY 927 GTTCGCTATTGACAAACTAGTAAATAACAGGTATACATGCGCCATGGCTCAGAA 986
 Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
 QY 987 ACCTTATACCTAGCTCTCTGAGGTGGGAGCATTTGGATCTATCATTTGGGCATCAATAT 1046
 Db 361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspIleTyrAspLeuGlyIleLysTyr 380
 QY 1047 TCGTT-----TAC 1054
 Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
 QY 1055 ATCAACCCACCTGTAGAGAGCTTTTGGCGTGTCTCTAAATAGCTGGCATGTCATT 1114
 Db 401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420
 QY 1115 AGGAATGTT 1123
 Db 421 ArgAsnVal 423

RESULT 5
 AA014733
 ID AA014733 standard; Protein; 423 AA.
 AC AA014733;
 XX
 DT 25-MAR-2003 (updated)
 DT 08-MAY-1997 (first entry)
 XX
 DE Human plasma carboxypeptidase B.
 XX
 KW Human, plasma carboxypeptidase B; PCPB; haemostatic regulation;
 KW plasma; plasminogen.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT Protein /note= "Signal peptide"
 FT Protein 23..423
 FT Cleavage-site 114 /note= "Mature PCPB"
 FT /note= "Potential clip site for activation of PCPB as a carboxypeptidase"
 FT Active-site 181 /note= "Expected to be involved in catalytic activity"
 FT Binding-site 183 /note= "Expected to be involved in substrate binding"
 FT Active-site 184 /note= "Expected to be involved in catalytic activity"
 FT Binding-site 256 /note= "Expected to be involved in substrate binding"

FT Binding-site 257 /note= "Expected to be involved in substrate binding"
 FT Active-site 310 /note= "Expected to be involved in catalytic activity"
 FT Binding-site 312 /note= "Expected to be involved in substrate binding"
 FT Active-site 369 /note= "Expected to be involved in catalytic activity"
 FT Binding-site 370 /note= "Expected to be involved in substrate binding, thought to determine specificity of PCPB as a carboxypeptidase B"
 FT Binding-site 371 /note= "Expected to be involved in substrate binding"
 FT Binding-site 394 /note= "Expected to be involved in substrate binding"
 FT
 XX US5593674-A.
 XX 14-JAN-1997.
 XX 27-APR-1995; 95US-0430787.
 XX 01-FEB-1991; 91US-0649591.
 PR 14-OCT-1992; 92US-0959944.
 PR 15-DEC-1993; 93US-0167727.
 PR 19-JUL-1994; 94US-0277540.
 PR 27-APR-1995; 95US-0430787.
 XX
 PA (GETH) GENENTECH INC.
 XX
 XX Drayna DT, Eaton DL;
 XX
 DR WPI; 1997-099413/09.
 DR N-PSDB; AAT62846.
 XX
 PT Using human plasma carboxypeptidase B in blood coagulation - is
 PT functionally related to carboxypeptidase A and pancreas
 PT carboxypeptidase B
 XX
 PS Example 2; Column 37-42; 39pp; English.
 XX
 CC This sequence represents human plasma carboxypeptidase B (PCPB) which
 CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
 CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
 CC purified from human plasma or by transformed cell culture by
 CC extraction using plasminogen bound to a solid phase.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 423 AA;
 Alignment Scores:
 Pred. No.: 3.24e-194 Length: 423
 Score: 1885.00 Matches: 365
 Percent Similarity: 86.52% Conservative: 1
 Best Local Similarity: 86.29% Mismatches: 3
 Query Match: 67.88% Indels: 55
 DB: 18 Gaps: 2
 US-09-980-881A-1 (1-1573) x AA014733 (1-423)

QY 18 ATGAAGCTTGGAGCCTTCAGTCCCTGTACCCATTGTTCTCTGTGAGCAGCATGTC 77
 Db 1 MetLeuLysCysSerLeuAlaValLeuValProIleValPheCysGluGlnHisVal 20
 QY 78 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGCTGTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
 Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
 QY 138 GTTCTACAGAACTTACTCAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
 Db 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

```
QY 198 CTTATTGCTGAAGAAAACAGTCCATTCTTTTGTAAATGCATCTGATGTCGACATGTC 257
Db 61 LeuileVallylsylsGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 258 AAAGCCCATTTAAATGAGCGGAATTCACAGACAGTCCAGTCTCTGCTGGCAGACGTGGAAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY 318 CTTATTCAACAGCAGATTTCACAGACACAGTCCAGCCCGAGCCCTCCGATCGTACTAT 377
Db 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
QY 378 GAACAGTATCACTCACTAAATCTATCTCTGATAGATAGATTAATTAACGTGAGGCAT 437
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 438 CCTGATATGCTTCAAAAATCCACATGGATCTCTATTTCAGAAAGTACCCACACTATGTT 497
Db 141 ProAspMetLeuThrIlylleHisIleGlySerSerPheGluIlySerProLeuTyrVal 160
QY 498 TTAAGGTTTTCTGAAAGAACAAACAGCCCAAAATGCCATATCGATTTGACTGTGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 558 CATGCCAGAAATGGATCTCTCTGCTTCTCTGCTTCTGCTGTTTCATAGGCCAT----- 608
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 608 ----- 608
Db 201 PheTyrGlyIlelleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 609 -----AATCGAATGTGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 627 AACGGTCTTTCTATGCGAACAATCATTTGATCGGAACAGACCTGGAATAGCAATTTGTC 686
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 687 TCAAAACACTGGTGTAGGAGGTGCATCCAGTCTCTCATGCTCGGAACCTACTGGA 746
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
QY 747 CTTTATCCTGAGTACAGAACAGAGTGAAGCAGTGTCTAGTCTTCTGAGAAGAAATATC 806
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 807 AACGAGATTAAAGCATATCAGCATCATCTATCTACTCCAGCATATAGTGTTCATAT 866
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 867 TCCTATACAGAACTAAAGCAAGACATGAGAACTGTCTAGTAGCCAGTGAAGCA 926
Db 321 SerTyrThrArgSerTyrSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
QY 927 GTTCGTCTATTGCAAAACTAGTAAAAATACCAAGGTATACACATGCGCCATGGCTGAGAA 986
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 987 ACCTTATACCTAGCTCTCGAGGTGGGACGATTTGATCTATGATTTGGGCATCAATAT 1046
Db 361 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1047 TCGTT-----TAC 1054
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1055 ATCAAAACCCACTGTAGAGAGCTTTTCCGCTGTCTCTAAATAGCTTGGCATGTCATT 1114
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLysIleAlaTrpHisValIle 420
QY 1115 AGGAATGTT 1123
```

```
Db 421 ArgAsnVal 423
RESULT 6
AAW92270
ID AAW92270 standard; Protein; 423 AA.
XX AAW92270;
XX 28-APR-1999 (first entry)
XX Human plasma carboxypeptidase B (PCPB) thr147.
XX Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
XX polymorphism detection; thrombotic disease.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..22
XX /note= "signal peptide"
XX Protein 23..423
XX /note= "mature PCPBthr147"
XX WO9855645-A1.
XX PD 10-DEC-1998.
XX PF 02-JUN-1998; 98WO-EP03244.
XX PR 03-JUN-1997; 97US-0869057.
XX (SCHD ) SCHERING AG.
XX Morser MJ, Nagashima M;
XX WPI; 1999-045800/04.
XX N-PSDB; AAW74302.
XX Detecting new polymorphism of human plasma carboxypeptidase B -
XX comprises Alanine or Threonine at position 147 of protein by DNA or
XX protein analysis, useful to detect risk of thrombotic disease in
XX humans
XX Disclosure; Page 25-26; 35pp; English.
XX This sequence is the human plasma carboxypeptidase B (PCPB) mutant
XX hPCPBthr147. The invention relates to a method for determining the
XX presence of DNA or protein polymorphisms of PCPB in human subjects, which
XX comprises obtaining a prepared tissue or blood sample and determining the
XX presence of DNA coding for naturally occurring polymorphs of the protein
XX containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
XX respectively). Determination of the relative distribution of the PCPB
XX polymorphs in a patient's blood by genetic or protein analysis by the
XX methods is useful to determine the risk of thrombotic disease in humans.
XX Such assessments may be made by accumulating information concerning the
XX relative distribution of the different polymorphs within the general
XX population compared with populations known to be at risk and establishing
XX a PCPB polymorph profile for at-risk patients.
XX Sequence 423 AA;
Alignment Scores:
Pred. No.: 3,24e-194 Length: 423
Score: 1885.00 Matches: 365
Percent Similarity: 86.52% Conservative: 1
Best Local Similarity: 86.29% Mismatches: 3
Query Match: 67.88% Indels: 55
DB: 20 Gaps: 2
US-09-980-881A-1 (1-1573) x AAW92270 (1-423)
QY 18 ATGAAGCTTTGCAGCCTTGCAAGTCCTTGATCCCATTTGTTCTTCTGTGACGACATGTC 77
```


Db 1 MetLysLeuCySerLeuAlaValLeuValProileValLeuPheCyGluGlnHisVal 20
QY 78 TTCCGGTTCCAGAGTGGCAAGTTCTAGCTCTCTAGAACCTCTAGCAAGTTCAA 137
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProargThrSerArgGlnValGln 40
QY 138 GTTCTACAGAACTTACTACAACATATGAGATTCTCTCGCAGCCGGTAAACAGCTGAC 197
Db 41 ValLeuGlnAsnLeuThrThrThrTyGluLeuValLeuTrpGlnProValThrAlaAsp 60
QY 198 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 257
Db 61 LeuileValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 80
QY 258 AAGCCCATTTAATGTGAGCGGAATTCATGCGAGTCTCTGCGCAGAGCTGGAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
QY 318 CTTATTCAACAGCAGATTTCACACGACACAGTCAAGCCCGAGCTCCGCATCTACTAT 377
Db 101 LeuileGlnGlnLilSerAsnAspThrValSerProargAlaSerAlaSerTyTy 120
QY 378 GAACAGTATCACTCACTAAATCTATTCTTGGATAGAATTATTAACCTGAGAGCAT 437
Db 121 GluGlnTyHisSerLeuAsnGluIleTySerTrpIleGluPheIleThrGluArgHis 140
QY 438 CTTGATATGCTTCAAAAAATCCCATTTGGATCTTCATTTGAGAAGTACCCACTATGTT 497
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeuTyVal 160
QY 498 TTAAGGTTTCTGAAAGAACAAACAGCAGCCAAAATGCCATATGATTGATGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCyGlyIle 180
QY 558 CATGCCAGAGAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGCCCAT----- 608
Db 181 HisAlaArgGluTrpIleSerProAlaPheCySerLeuTrpPheIleGlyHisIleThrGln 200
QY 608 ----- 608
Db 201 PheTyGlyIleIleGlyGlnTyThrAsnLeuLeuArgLeuValAspPheTyValMet 220
QY 609 -----AATCGAATGCGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyAspTySerTrpLysLysAsnArgMetTrpArgLys 240
QY 627 AACCGTTCTTTCTATGCGAACAAATCATTTGCATCGGAACAGACCTGATAGCACTTTGTC 686
Db 241 AsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 687 TCCAAAACATCGTGTGAGGAGGTGCATCGAGTTCTCTGCTCGGAAACCTACTGTGGA 746
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyCysGly 280
QY 747 CTTTATCTCTGAGTACAGACAGAGTCAAGCGAGTGGCTAGTTCTTGTGAGAGAAATATC 806
Db 281 LeuTyProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 807 AACAGATTAAGACATACATCAGCATGCATTCTACTCTCCAGCATATAGTGTTCATAT 866
Db 301 AsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleValPheProTy 320
QY 867 TCCTATACAGAACTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 926
Db 321 SerTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
QY 927 GTTCGTCTATTGACAAACTAGTAAATAACAGGATATACATATGCCATGCTCAGAA 986
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisGlyHisGlySerGlu 360
QY 987 ACCTTATACCTAGCTCTCGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAATAT 1046

Db 361 ThrLeuTyLeuAlaProGlyGlyGlyAspAspTrpIleTyTrAspLeuGlyIleLysTy 380
QY 1047 TCCTT-----TAC 1054
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyGlyPheLeuLeuProGluArgTy 400
QY 1055 ATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAAATAGCTTGGCATGTCA 1114
Db 401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420
QY 1115 AGGAATGTT 1123
Db 421 ArgAsnVal 423
RESULT 7
AAR36273
ID AAR36273 standard; Protein; 423 AA.
XX
AC AAR36273;
XX
DT 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22 /label= signal_peptide
FT Cleavage-site 114 /note= "putative site of cleavage by trypsin to activate PCPB as a carboxypeptidase"
FT Active-site 181 /note= "expected to be involved in catalytic activity"
FT Binding-site 183 /note= "expected to be involved in substrate binding"
FT Active-site 184 /note= "expected to be involved in catalytic activity"
FT Binding-site 256..257 /note= "expected to be involved in substrate binding"
FT Active-site 310 /note= "expected to be involved in catalytic activity"
FT Binding-site 312 /note= "expected to be involved in substrate binding"
FT Active-site 369 /note= "expected to be involved in catalytic activity"
FT Binding-site 370..371 /note= "expected to be involved in substrate binding; Asp 370 is the residue which determines specificity of the PCPB as a carboxypeptidase B"
FT Binding-site 394 /note= "expected to be involved in substrate binding"
XX
PN US5206161-A.
XX
PD 27-APR-1993.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
XX
PA (GETH) GENENTECH INC.

XX Drayna DT, Baton DL;
XX WPI; 1993-151724/18.
DR N-PSDB; AAQ41001.
XX New human plasma carboxypeptidase B - used as haemostatic
PT regulator for clotting blood, partic. for treating blood clotting
PT disorders, e.g. haemophilia
XX Claim 2; Fig 4; 40pp; English.
XX This amino acid sequence was deduced from a clone isolated from a
CC human liver cDNA library. There is about 40% sequence identity
CC between the deduced preproPCPB amino acid sequence and pro-human
CC mast cell carboxypeptidase A and between preproPCPB and prepro-rat
CC carboxypeptidase B. Human PCPB has the same substrate binding sites
CC as, and shares 6 cysteine residues which form 3 intramolecular
CC disulphide bonds with, bovine and rat carboxypeptidase B. The
CC presence of the same amino acid (Asp) at position 348 of the mature
CC protein) at the region in carboxypeptidases that determines
CC substrate specificity, suggests that PCPB represents a plasma-
CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC tPA of plasminogen to plasmin in the presence of fibrinogen.
XX SQ Sequence 423 AA;
Alignment Scores:
Pred. No.: 1,12e-193 Length: 423
Score: 1880.00 Matches: 364
Percent Similarity: 86.29% Conservative: 1
Best Local Similarity: 86.05% Mismatches: 4
Query Match: 67.70% Indels: 55
DB: 14 Gaps: 2
US-09-980-881A-1 (1-1573) x AAR36273 (1-423)
QY 18 ATGAAGCTTTGACGCTTGTGAGTCTTGTACCATTTCTTCTGTGTGAGCAGCATGTC 77
DB 1 MetLysLeuCySerLeuAlaValLeuValProileValLeuPheCyGluGlnHisVal 20
QY 78 TTCGCTTCAGAGTCCAGTCTAGCTCTTCTAGACCTTAGCGAAGTCAA 137
DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 138 GTTCTACAGAACTTACTCAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 197
DB 41 ValLeuGlnAsnLeuThrThrThrGluileValLeuTrpGlnProValThrAlaAsp 60
QY 198 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGTATGTCGAATGTG 257
DB 61 LeuileValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 258 AAGCCCATTTAATGTGACGGAAATTCATGAGTGTCTTGTGGCAGACGTGGAGAT 317
DB 81 LysAlaHisLeuAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
QY 318 CTTATTCAACAGCAGATTTCCACACACACAGTCCAGCCCGAGCTCCGCTCATCTAT 377
DB 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyTr 120
QY 378 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAATTATTAACGTGAGGCGAT 437
DB 121 GluGlnTyHisSerLeuAsnGluileTySerTrpIleGluPheIleThrGluArgHis 140
QY 438 CTTGATATGCTTCAAAAATCCCATTTGATTCCTCATTTGAGAGTACCCACTTATGTT 497
DB 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeuTyTrVal 160
QY 498 TTTAAGCTTTCTGAAAGAACCAACACAGCCAAAATGCCATATGCTATGCTGGAATC 557
DB 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCyGlyIle 180
558 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGGTTCATAGGCCAT----- 608
181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
608 ----- 608
201 PheTyGlyIleIleGlyGlnTyThrAsnLeuLeuArgLeuValAspPheTyValMet 220
609 -----AATCGAATGTGGAGAAAG 626
221 ProValValAsnValAspGlyTyAspTySerTrpLysLysAsnArgMetTrpArgLys 240
627 AACCGTCTTCTTATCGGAACATCATTCATCGGACAGACCTGATAGCACTTTGTC 686
241 AsnArgSerPheTyTrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
687 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATCTGCTCGGAAACCTACTGTGA 746
261 SerLysHisTrpCysGluGluGlyAlaCysSerSerCysSerGluThrTyCysGly 280
747 CTTTATCTGAGTCAGAACCCAGAGTGAAGGCTGCTAGTTCTTGTGAGAGAAATATC 806
281 LeuTyTrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
807 AACGATTAAAGCATACATCAGCATGCTATCTACTCCAGCATATAGTGTTCATAT 866
301 AsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleValPheProTy 320
867 TCCTATACAGAGTAAGAAAGAACCATGAGGAACCTGTCTAGTACGACGTGAAGCA 926
321 SerTyTrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
927 GTTCGTGCTATTGACAAACTAGTAAAAATACACAGGTATACACATGGCCATGGCTCAGAA 986
341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisGlyHisGlySerGlu 360
987 ACCTTATACCTAGCTCTGAGGTGGGAGGATGATGATCATGATTTGGCATCAATAT 1046
361 ThrLeuTyLeuAlaProGlyGlyGlyAspTrpIleTyAspLeuGlyIleLysTy 380
1047 TCGTT-----TAC 1054
381 SerPheThrIleGluLeuArgAspThrGlyTyGlyPheLeuLeuProGluArgTy 400
1055 ATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAAATAGCTTGGCATGTCA 1114
401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420
1115 AGGAATGTT 1123
421 ArgAsnVal 423
RESULT 8
AAB11458
ID AAB11458 standard; Protein; 338 AA.
XX
XX AAB11458;
XX
XX 01-MAR-2001 (first entry)
XX
XX Human brain carboxypeptidase B protein SEQ ID NO 3.
XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; antialzheimers; neurotropic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma.
XX Homo sapiens.
XX
XX WO2000066717-A1.
XX
XX 09-NOV-2000.

XX 01-MAY-2000; 2000WO-JP02878.
 XX PF
 XX PR 30-APR-1999; 99JP-0125169.
 XX PA (MATS/) MATSUMOTO A.
 XX PI Matsumoto A;
 XX DR WPI; 2000-687534/67.
 XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX PS Disclosure; Page 71-73; 84pp; Japanese.
 XX CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimer's,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 XX SQ Sequence 338 AA;
 Alignment Scores:
 Pred. No.: 1.82e-183 Length: 338
 Score: 1785.00 Matches: 335
 Percent Similarity: 99.41% Conservative: 1
 Best Local Similarity: 99.11% Mismatches: 2
 Query Match: 64.28% Indels: 0
 DB: 21 Gaps: 0
 US-09-980-881A-1 (1-1573) x AAB11458 (1-338)
 QY 84 TTCAGAGTGGCCAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAGTTCTA 143
 Db 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 QY 144 CAGAACTCTTACTCAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTATT 203
 Db 21 GlnAsnLeuThrThrThrGluValLeuValLeuValLeuValLeuValLeuValLeu 40
 QY 204 GTGAAGAAAAACAAAGTCCATTTTGTAAATGCATCTGATGTGCGAATGTGAAAGCC 263
 Db 41 ValLysLysLysGlnValHisPheValAsnAlaSerAspValAspValLysAla 60
 QY 264 CATTTAAATGTAGCCGAATTCATGTAGTGTCTTGTGGCAGACGTGGAAGATCTATT 323
 Db 61 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
 QY 324 CAACAGCAGATTTCCAAACGACACAGTCAGCCCGCCGAGCTCCGATCGTACTATGACAG 383
 Db 81 GlnGlnGlnLeuSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 QY 384 TATCACTCACTAAATGAATCTATTCTTGGATGAATTTTATACTGAGAGCGCATCTCTGAT 443
 Db 101 TyrHisSerLeuAsnGluLeuTyrSerTrpLeuGluPheLeuThrGluArgHisProAsp 120
 QY 444 ATGCTTCAAAAATCCCATTTGGATTCATTTTGGAGAAAGTACCCATCTATGTTTAAAG 503
 Db 121 MetLeuThrLysLeuHisLeuGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
 QY 504 GTTCTCGAAAGAACAAACAGCCAAATGCCATATGATTCGATTCGATTCGAAATCCATGCC 563
 Db 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaLeuTrpLeuAspCysGlyLeuHisAla 160
 QY 564 AGAGAATGGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 623

Db 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
 QY 624 AAGAACGGTCTTTCTATCGAACAAATTCATTCGATCGAACACCTGATAGCACTTTT 683
 Db 181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
 QY 684 GTCTCCAAACACATGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743
 Db 201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 220
 QY 744 GGACTTTATCTCAGTCAGAACACAGAGTGAAGCGAGTGGCTAGTTCCTTTGAGAAGAAAT 803
 Db 221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 240
 QY 804 ATCAACCAAGTTAAAGCATACATCAGCATCGCATTCATCTCCAGCATATAGTCTTCCA 863
 Db 241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
 QY 864 TATTCTTATACACGAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAA 923
 Db 261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
 QY 924 GCAGTTCTGCTATTACACAAAATCTACTAAAATACCAAGTATACATGCGCATGCTCA 983
 Db 281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
 QY 984 GAAACCTTATACCTAGCTCTCTGAGGTGGGAGCATTTGATCTATCATTTGGGCATCAA 1043
 Db 301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyLeuLys 320
 QY 1044 TATTCTGTTTATCATCAACCCACCTGTAGAGAAGCTTTTGGCGCTGCTCTTAAAA 1097
 Db 321 TyrSerPheThrSerAsnProValGluLysLeuLeuProLeuSerLeuLys 338
 RESULT 9
 AAB11459
 ID AAB11459 standard; Protein; 246 AA.
 AC AAB11459;
 XX 01-MAR-2001 (first entry)
 DT Human brain carboxypeptidase B protein SEQ ID NO 4.
 XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 DE treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.
 XX Homo sapiens.
 OS WO200066717-A1.
 PN 09-NOV-2000.
 PD 01-MAY-2000; 2000WO-JP02878.
 PF 30-APR-1999; 99JP-0125169.
 PR (MATS/) MATSUMOTO A.
 PA Matsumoto A;
 XX WPI; 2000-687534/67.
 DR Human brain carboxypeptidase B isolated from the hippocampus useful for
 PT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX PS Disclosure; Page 74-75; 84pp; Japanese.
 XX CC This invention describes a novel protein with peptidase activity

CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimer's,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
xx
xx Sequence 246 AA:

Alignment Scores:		
Pred. No.:	7, 22e-134	246
Score:	1325.00	243
Percent Similarity:	99.19%	Conservative: 1
Best Local Similarity:	98.78%	Mismatches: 2
Query Match:	47.71%	Indels: 0
DB:	21	Gaps: 0

US-09-980-881A-1 (1-1573) x AAB11459 (1-246)

360	Qy	GCCTCCGCATCGTATGATGAACAGATCACCTCAATAAAGAAATCTATATCTTCGTGATGAA	419
1	Db	AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu	20
420	Qy	TTTATAACTGAGAGGCATCTGATATGCTTACAAAATCCACATTGGATCCTCATTTGAG	479
21	Db	PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu	40
480	Qy	AGATACCCACTCTATGTTTAAAGGTTCTCGAAAAAGAACAAACAGCCAAAATATGCCATA	539
41	Db	LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle	60
540	Qy	TGGATTGACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTTTGTGTGTC	599
61	Db	TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe	80
600	Qy	ATAGGCCATAATCGAATGGGAGGAAGAACCGTTCTTTCTATGCGAACATCATTCGATC	659
81	Db	IleGlyHisAsnArgWecTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle	100
660	Qy	GGAAACAGACTCAATAGCAACATTTCTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT	719
101	Db	GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer	120
720	Qy	TCCTCATGCTCGAAACCTACTGTGGAATTTATCTCTGAGTCAGAACACAGAAGTGAAGGCA	779
121	Db	SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla	140
780	Qy	GTGGCTAGTTCTTCGAGGAAGAATATCAACAGATTAAGCATATACATCAGCATGCATCCA	839
141	Db	ValAlaSerPheLeuArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer	160
840	Qy	TACTCCAGCATATAGTGTTCATATTCCTATACAGAACTAAAGCAAGACCATGAG	899
161	Db	TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu	180
900	Qy	GAATCTCTCTAGTCGACGATGAGCAGTTCGTGCTATTTCACAAAATAGTAAAAATACC	959
181	Db	GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr	200
960	Qy	AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGAGGTGGGGACGAT	1011
201	Db	ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp	220
1020	Qy	TGGATCTATGATTGGGCATCAAAATATTCGTTTATCATCAAAACCCACCTGTAGAGAAGCTT	1071
221	Db	TrpIleTyrAspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeu	240
1080	Qy	TTGCCGCTGCTCTAAAA	1097
241	Db	LeuProLeuSerLeuLys	246

RESULT 10

AAB58129
ID AAB58129 standard; Protein; 211 AA.

12
XX
AC

AC
XX
DT

XX DE Lung cancer associated polypeptide sequence SEQ ID 467.

Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotropic; antinefactive; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reprodu
 KW proliferative disorder; wound healing; infectious disease.
 KW lung cancer associated polypeptide sequence **seq** **15** **407**.

XX
OS Homo sapiens.

XX PN WO2000055180-A2

FN
XX
PD

FD 21-SEP-2000.
XX
DE 08-MAR-2000. 200000-11005010

PF 08-MAR-2000; 2000WC-US05918.
XX
DE 13 MAR 1000 0011C 012423Z

PR 12-MAR-1999; 990S-0124270.
XX
D

PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.

XX
PI Ruben SM;

XX
DR
WPI; 2000-587514/55.

DR N-PSDB; AAF18005.
XX

PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -

PS Claim 11: Page 944-945: 1425pp: English.

XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.

XX Sequence 211 AA:

Alignment Scores:

Pred. No.:	7,96e-87	Length:	211
Score:	899.00	Matches:	173
Percent Similarity:	83.33%	Conservative:	2
Best Local Similarity:	82.38%	Mismatches:	9
Query Match:	32.01%	Indels:	27
DB:	21	Gaps:	3

US-09-980-881A-1 (1-1573) x AAB58129 (1-211)

Qy 570 TGGATCTCTCCTGCTTCTCTGCTTGTGTTGTTTCATA

570 TGGATCTCTCCTGCTTCTGCTTGTGGTTCATA-----GGC 605

Db 3 TrpIleSer---MetLeuCysArgTrpLeuMetTrpMetValMetAsnTyrSerTrpLys 21
Qy 606 CATAATCGAATGTGGAGAAACCGTCTTCTATCGCAACATATTCATCGGACAA 665
Db 22 LysAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThr 41
Qy 666 GACCTGATAGCAACTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCA 725
Db 42 AspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSer 61
Qy 726 TGCTCGGAACCTACTGTGAGTTCCTGAGTCAGACACAGAGTGAAGCGGTGGCT 785
Db 62 CysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAla 81
Qy 786 AGTTTCTTGAGAAATATCAACACAGATTAACATACATACATCAGCATGCATTCATCTCC 845
Db 82 SerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSer 101
Qy 846 CAGCATATAGTGTTCATATCTTATACAGAGTAAAGCAAGACCATGAGGAAGTCTG 905
Db 102 GlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 121
Qy 906 TCCTAGTAGCAGTGAAGAGTTCGTCTATTGACAAAATAGTAAATAACAGATAT 965
Db 122 SerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyr 141
Qy 966 ACATCGCCATGCTCAGAAACCTTATACCTAGCTCTCGAGTGGGAGCATGGATC 1025
Db 142 ThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspSerTrpIle 161
Qy 1026 TATGATTGGGCATCAAAATATTCCTT-----1051
Db 162 TyrAspLeuGlyIleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGly 181
Qy 1052 -----TACATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCT 1093
Db 182 PheLeuLeuProGluArgTyrIleLysProThrCysArgGluAlaPheAlaValSer 201
Qy 1094 AAAATAGCTGGCATGTCATTAGGAATGTT 1123
Db 202 LysIleAlaThrHisValIleArgAsnVal 211
RESULT 11
ABB99688
ID ABB99688 standard; Protein; 417 AA.
AC ABB99688;
XX 28-MAR-2003 (first entry)
DT
DE
XX Amino acid sequence of human zinc carboxypeptidase B precursor.
XX Human; zinc carboxypeptidase B-like protein; cancer; CNS disorder;
KW chronic obstructive pulmonary disease; COPD; haematological disorder;
KW zinc carboxypeptidase B.
XX Homo sapiens.
OS
XX WO200295018-A1.
PN
XX 28-NOV-2002.
PD
XX 16-MAY-2002; 2002WO-EP05392.
PF
XX 22-MAY-2001; 2001US-292329P.
PR
XX (FARB) BAYER AG.
PA
XX Smolyar A;
XX WPI; 2003-120801/11.
XX
XX New polynucleotide encoding a zinc carboxypeptidase B-like protein

PT polypeptide, useful for treating diseases associated with
PT carboxypeptidase dysfunction, e.g. cancer, CNS disorder, COPD or
PT haematological disorder
XX
XX Disclosure; Fig 3; 113pp; English.

XX The present sequence represents a zinc carboxypeptidase B precursor.

CC The specification describes a human zinc carboxypeptidase B-like
CC protein. The zinc carboxypeptidase B-like protein is used for treating
CC diseases, such as cancer, a CNS disorder, chronic obstructive pulmonary
CC disease (COPD) or haematological disorders. The zinc carboxypeptidase
CC B-like protein may also be used to identify compounds which may act
CC as activators or inhibitors at the enzyme's active site, to raise
CC specific antibodies which can block the enzyme and effectively reduce
CC its activity, as a bait protein in a two-hybrid or three-hybrid assay
CC to identify other proteins which bind to or interact with the human
CC zinc carboxypeptidase-like protein polypeptide and modulate its
CC activity, and for immunization of mammals.

XX Sequence 417 AA;

Alignment Scores:

Pred. No.:	3,81e-61	Length:	417
Score:	652.50	Matches:	142
Percent Similarity:	52.13%	Conservative:	78
Best Local Similarity:	33.65%	Mismatches:	141
Query Match:	23.50%	Indels:	62
DB:	24	Gaps:	8

US-09-980-881A-1 (1-1573) x ABB99688 (1-417)

Qy 24 CTTTGCAGCTTGCAGCTCTGTACCATTTCTCTCTGTGTGAGCAGCATGTCTTCGCG 83

Db 1 MetLeuAlaLeuLeuValThrValAlaLeuAlaSerAlaHisGlyGlyGlu 20

Qy 84 ---TTCAGAGTGGCCAAAGTCTTAGCTGCTCTCTCTAGAACCTCTAGGCAAGTTCAGTT 140

Db 21 HisPheGluGlyLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy 141 CTACAGATCTTACTACACATATGAGATGTTCTCTGCGAGCCGGTACACAGTGCCTT 200

Db 41 IleArgLysLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy 201 ATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATCTCGACATGTGAA 260

Db 61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy 261 GCCCATTTAAATGTGAGCGGAATTCATGCAGTGTCTTGTGCGACAGCTGGAAGATCTT 320

Db 81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy 321 ATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCGGAGCTCCGCTACTATGAA 380

Db 101 ValGluAlaGlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGlu 117

Qy 381 CAGTATCACTCAATAATGAATCTATTCTTGATAGATTTATACTGAGGCGATCTT 440

Db 118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy 441 GATATGCTTACAAAAATCCACATTTGATGCTCATTTTGAGAAGTACCCTATGTTTAA 500

Db 138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy 501 AAGGTTTTCGAAAGAAACAAACAGCCAAATATGCGCATATGATGATGATGATGAT 560

Db 158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy 561 GCCAGAGATGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602

Db 177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy 602 -----602

```
Db 197 TyrGlyArgGluLeuGlnValThrGluLeuLeuAsnLysLeuAspPheTyrValLeuPro 216
Qy 603 -----GCCCAT-----AATCGAATGTCGAGAAAGAAC 629
Db 217 ValLeuAsnLeuAspGlyTyrIleTyrThrThrLysSerArgPheTyrArgLysThr 236
Qy 630 CGTCTCTTCATCGAACAATCATTCGATCGGACAGACCTGAATAGCACTTTGCTCC 689
Db 237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255
Qy 690 AAACACTGGTGTGAGGAAGGTGATCCAGTCTCTCGTCCGAAACCTACTGTGGACTT 749
Db 256 AlsGlyTyrCysGluLeuGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275
Qy 750 TATCTGAGTCAGAACAGAGTGAAGCGAGTGTCTCTGTGAGAGAAATATCAAC 809
Db 276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
Qy 810 CAGATTAAACATACATCAGCATGCTATCTACTCCAGCATATAGTGTTCCTATATCC 869
Db 296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315
Qy 870 TATACACGAAGTAAAGCAAGACCATGAGCACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
Db 316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLysAlaThrVal 335
Qy 930 CGTGTCTATTGACAAAACACTAGTAAATACCAGGTATACATGCGCCATGCGCTCAGAAAC 989
Db 336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354
Qy 990 TTATACCTAGCTCTGAGGTGGGAGCATGATGATCTATGATTTGGGCATCAATATCG 1049
Db 355 IleTyrProAlaAlaGlyGlySerAspAspTyrAlaTyrAspGlnGlyIleArgTyrSer 374
Qy 1050 TTTATCATC----- 1057
Db 375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394
Qy 1058 AAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAAATAGCTTGGCATGTCAATAGG 1117
Db 395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGlu 414
Qy 1118 ATGTTT 1123
Db 415 HisLeu 416

RESULT 12
AAW06172
XX 12-FEB-1997 (first entry)
AC AAW06172;
XX Human pancreatic carboxypeptidase B.
DT 12-FEB-1997 (first entry)
XX Human pancreatic carboxypeptidase B.
DE ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW mustard-ribonuclease; antibody directed enzyme prodrug therapy;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Pg; F(ab')2; PeIB; leader; human carboxypeptidase B.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..13
FT /note= "partial enzyme secretion leader sequence;
FT (pre-sequence)"
FT Peptide 14..108
FT /label= pro-sequence
FT Protein 109..415
FT /label= mature_enzyme
FT Misc-difference 243
```

```
FT FT /note= "not observed in other published HPCB
FT FT sequences"
FT Active-site 361
FT /note= "Asp253 characteristic of B type specificity"
PN WO9620011-A1.
XX 04-JUL-1996.
XX 21-DEC-1995; 95WO-GB02991.
XX 16-AUG-1995; 95GB-0016910.
XX 23-DEC-1994; 94GB-0026192.
XX (ZENE ) ZENECA LTD.
XX Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
XX Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
XX Tarragona-Fiol A, Taylorson CJ;
XX WPI; 1996-321650/32.
XX N-PSDB; AAT42494.
XX Two component system for anti-tumour therapy - comprising targeting
XX moiety linked to mutated enzyme which can transform an
XX anti-neoplastic prodrug
XX Reference Example 14; Page 131-132; 182pp; English.
XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX prodrug. The system is based on antibody directed enzyme prodrug therapy
XX (ADEPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
XX Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The present sequence is that of the native prepro-HCPB.
XX
SQ Sequence 415 AA;
Alignment Scores:
Pred. No.: 4,87e-61 Length: 415
Score: 651.50 Matches: 142
Percent Similarity: 52.03% Conservative: 76
Best Local Similarity: 33.89% Mismatches: 140
Query Match: 23.46% Indels: 62
DB: 17 Gaps: 8
US-09-980-881A-1 (1-1573) x AAW06172 (1-415)
Qy 33 CTTGCAGTCTCTTACCCATTGTCTCTGTGAGCAGCATGCTTTCGCG---TTCCAG 89
Db 2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisGlyGlyGluHisPheGlu 21
Qy 90 AGTGGCCAAAGTTCTAGCTGCTCTTCTCTAGAACCTCTTAGCAAGTTCAAGTTCTACAGAT 149
Db 22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41
Qy 150 CTTTACTACACATATGAGATTGTTCTCTGCAGCCGTAACAGCTGACCTTATTGTGAAG 209
Db 42 LeuAlaSerThrThrGlnIleAspPheThrLysValLeuIleSerAsnLeuArgAsnValGluAla 61
Qy 210 AAAAAACAAGTCCATTTTGTGTAATGTCATCTGTGTCACAAATGTGAAGGCCCATTTA 269
Db 62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81
Qy 270 AATGTGAGCGGAATTCATGCACTGCTGCTGGCAGACGTGAAGATCTTATTCAACAG 329
Db 82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValGluAla 101
Qy 330 CAGATTTCACACACACAGTCAGCCCCCGAGCTCCGCATCGTACTATGACAGATATCAC 389
Db 102 GlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118
```



```

Db      119 LysTrpGluThrIleGluAlaThrThrGlnValAlaThrGluAsnProAlaLeuIle 138
Qy      450 AAAAAATCCACATTCCTCATTTTGAAGPACCCACTCTATGTTTAAAGTTTCT 509
Db      139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal-- 157
Qy      510 GGAAGAACAACACAGCAAAATGCCATATGATTCATGATTCGATTCGATTCGACAGAA 569
Db      158 GlyLysAlaGlyGlnAenLysProAlaIlePheMetAspCysGlyPheHisAlaAaGlu 177
Qy      570 TGCATCTCTCTGCTTCTGCTGCTGCTCATA----- 602
Db      178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197
Qy      602 ----- 602
Db      198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217
Qy      603 -----GGCCAT-----AATCGAATGTGGAGAAAGAACCGTCTCTTC 638
Db      218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237
Qy      639 TATCGAACAATCATTCGATCGAAGACAGCTGAATAGCAACTTGTCTCCAAACACTGG 698
Db      238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256
Qy      699 TGTGAGGAAGTCATCCAGTCTCTCATGTCGGAACCTACTGTGAGCTTTATCTCTGAG 758
Db      257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276
Qy      759 TCAGAACCAGAGTGAAGGAGTGGTGTAGTTCTTGTGAGAAGAATATCAACAGATTAA 818
Db      277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296
Qy      819 GCATACATCAGTCATTCATCTACTCCAGCATATAGTGTTCATATTCCTATACAGA 878
Db      297 AlaTyrLeuThrIleHisSerTyrSerGlnMetIleTyrProTyrSerTyrAlaTyr 316
Qy      879 AGTAAAGAACACCATGAGGAAGTCTCTCTAGTACAGTGAAGCAGTTCGTGCTATT 938
Db      317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336
Qy      939 GACAAACTAGTAAAAATACCAAGTATACATGCGCATGGCTCAGAAACCTTATACCTA 998
Db      337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355
Qy      999 GCTCTCGGAGTGGGACGATTCGATCTATGATTTGGGCATCAATATTCGTTTACATC- 1057
Db      356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375
Qy      1058 -----NAACCCACC 1066
Db      376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395
Qy      1067 TGTGAGAAGCTTTTGGCGCTGTCTCTAAATAGCTTGGCATGTCAATTAGGAATGTT 1123
Db      396 CysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHisLeu 414

RESULT 14
AAU84325
ID      AAU84325 standard; Protein; 417 AA.
XX
AC      AAU84325;
XX
DT      08-MAY-2002 (first entry)
XX
DE      Protein CPA3 differentially expressed in breast cancer tissue.
XX
KW      Human; diagnosis of breast cancer; endometrial cancer; breast tumour;
XX      MAI; mitotic activity index; cytostatic.
XX
OS      Homo sapiens.

```

```

XX      WO200210436-A2.
XX
XX      07-FEB-2002.
XX
XX      27-JUL-2001; 2001WO-US23642.
XX
XX      28-JUL-2000; 2000US-222093P.
XX
XX      (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX      (BAAK/) BAAK J.
XX
XX      Baak J, Mutter GL;
XX
XX      WPI; 2002-180084/23.
XX      N-PSDB; ABK35545.
XX
XX      Diagnosing breast cancer comprises determining expression of nucleic
XX      acid molecules or expression products that are differentially expressed
XX      in normal and malignant tissue -
XX
XX      Claim 37; Page 134-135; 219pp; English.
XX
XX      The present invention relates to a method for diagnosing breast cancer
XX      in a subject suspected of having endometrial cancer. The method
XX      comprises determining the expression of a set of human genes or
XX      expression products in an endometrial sample suspected of being
XX      cancerous. The human genes of the invention are differentially
XX      expressed in breast tumours characterised as high or low MAI (mitotic
XX      activity index). These sets of genes can be used to discriminate between
XX      high and low MAI breast tumours. The invention also provides DNA and
XX      protein microarrays for analysing the expression of the human genes and
XX      their protein products. The methods and arrays are useful for the
XX      diagnosis and prognosis of endometrial cancer, selecting and monitoring
XX      treatment regimes, and identification of compounds useful for the
XX      treatment of endometrial cancer. AAU84311-AAU84361 represent the human
XX      proteins of the invention that are differentially expressed in breast
XX      cancer tissue.
XX
XX      SQ      Sequence 417 AA;

```

```

Alignment Scores:
Pred. No.: 1,32e-60      Length: 417
Score: 647.50      Matches: 143
Percent Similarity: 52.46%      Conservative: 81
Best Local Similarity: 33.49%      Mismatches: 132
Query Match: 23.32%      Indels: 72
DB: 23      Gaps: 10
US-09-980-881A-1 (1-1573) x AAU84325 (1-417)

```

```

Qy      6 ATGCTGTGGGATGAAGCTTTGCAGCCTTGACGCTTGTACCCATTGTTCTTCTGT 65
Db      5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal----- 19
Qy      66 GAGCAGCATGCTCTTCGCTTCAGAGTTCAGAGTTCAGTCTCTCTCTCTAGAACCTCT 125
Db      20 -----ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34
Qy      126 AGGCAAGTTCAAGTTCTTACAGAACTTACTACACATATGAGATTTGTTCTTGGCAGCCG 185
Db      35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTyrPro 54
Qy      186 GTAACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGATCATCAT 245
Db      55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74
Qy      246 GTCGACAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGTCAGTGTCTTGTCTGGCA 305
Db      75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMethHisTyrGluIleLeuHis 94
Qy      306 GAGCTGGAAGATCTTATTCAACAGCAGATTTTC---AACGACACAGTCAGCCCGAGCC 362

```


Db 95 AspleuGlnGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114
 QY 363 TCGCATCGTACTATGAACAGTATCACTCAATAATCAATCAATCTATTCTCGATAGAAATTT 422
 Db 115 Ser-----TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131
 QY 423 ATAACCTGAGGCGATCCTGATGCTTACAAAATCCACATCCGATCCCTCATTTGGAAG 482
 Db 132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151
 QY 483 TACCACCTATGTTTAAAGGTTCTGGAAGAACAAACAGCCAAATGCGATATGG 542
 Db 152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170
 QY 543 ATTGACTGTGGAATCCATGCAGAGATGATCTCTCTGCTTCTGCTTGTGCTTCATA 602
 Db 171 MetAspCysGlyIleHisAlaArgLutrpValSerProAlaPheCysGlnTrpPheVal 190
 QY 602 ----- 602
 Db 191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210
 QY 603 -----GGCCAT-----AAT 611
 Db 211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230
 QY 612 CGAATGCGAGAGACCGCTCTTCTATCGAACAAATCATTCGATCGGAACAGACCTG 671
 Db 231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250
 QY 672 AATAGCAATCTTGTCTCCAAACACCTGTGTGAGGAGTGCATCCAGTCTCTCATGCTCG 731
 Db 251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269
 QY 732 GAAACCTACTGTGACTTTATCTGATCAGAACAGCAAGTGAAGCAGTGGCTAGTTTC 791
 Db 270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289
 QY 792 TTGAGAGAATAATACACAGATTAAGCATACATCAGCATGCATTCATCTATCTCCACCAT 851
 Db 290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309
 QY 852 ATAGTGTTCATATTCCTATACAGAAAGTAAAGCAACAGACCACTAGGAACTGTCCTA 911
 Db 310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329
 QY 912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAATAGTAAATACACAGGTATACAT 971
 Db 330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348
 QY 972 GGCCATGGCTCAGAACTTATACCTAGTCTCTGAGGTGGGAGCATTTGATCTATGAT 1031
 Db 349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368
 QY 1032 TTGGGCATCAATATTCGTTTAC----- 1054
 Db 369 LeuGlyIleHisIleThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388
 QY 1055 -----ATCAAAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAATA 1099
 Db 389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408
 QY 1100 GTTTGGCATGTCATGAAT 1120
 Db 409 AlaLysTyrIleLeuLysHis 415
 RESULT 15
 ID ABP97414 standard; Protein; 667 AA.
 AC ABP97414;
 XX
 DT 30-MAY-2003 (first entry)

XX Anti-human seminal plasma protein ScFv/hCPA fusion protein.
 DE Single chain antibody: ScFv; anti-human seminal plasma protein;
 XX monoclonal antibody E4B7; mouse; murine; human; carboxypeptidase A;
 KW hCPA; fusion protein; antibody-directed enzyme prodrug therapy; ADEPT;
 KW methotrexate-alpha-peptide prodrug; prostate cancer; cytostatic;
 KW gene; ds.
 XX Chimeric - Mus musculus.
 OS Chimeric - Homo sapiens.
 OS Synthetic.
 OS
 PH Key Location/Qualifiers
 FT Protein 1..245
 FT /label= E4B7 derived ScFv
 FT /note= "Anti-human seminal plasma protein single
 chain antibody derived from murine monoclonal
 antibody E4B7 VH and VL regions; targets
 fusion protein to prostate cancer cells"
 FT Region 1..122
 FT /note= "Corresponds to monoclonal antibody E4B7 VH (heavy
 chain variable region)"
 FT Region 123..137
 FT /label= ScFv linker
 FT Region 138..245
 FT /note= "Corresponds to monoclonal antibody E4B7 VL (light
 chain variable region)"
 FT Misc-difference 216
 FT /note= "Encoded by GAC"
 FT Region 245..251
 FT /label= ScFv/hCPA linker
 FT Misc-difference 247
 FT /note= "Encoded by AGC"
 FT Protein 253..667
 FT /label= hCPA
 FT /note= "Human carboxypeptidase A; activates methotrexate-
 alpha-peptide prodrug"
 XX WO2002100431-A1.
 PN 19-DEC-2002.
 XX 08-JUN-2001; 2001WO-CN00924.
 XX 08-JUN-2001; 2001WO-CN00924.
 PR (UYTA-) UNIV TANGDU HOSPITAL FORTH MILITARY MEDI.
 PA (HAOX/) HAO X.
 XX Hao X;
 PI WPI; 2003-156924/15.
 XX N-PSDB; AB275116.
 DR
 DR
 XX Pharmaceutical kits for use in the treatment of prostate cancer,
 PT comprise anti-human seminal plasma protein single-chain antibody/human
 PT carboxypeptidase fusion protein and a prodrug of
 PT methotrexate-alpha-peptides -
 XX Claim 3; Fig 4; 29pp; Chinese.
 XX The invention relates to a pharmaceutical kit for antibody-directed
 CC enzyme prodrug therapy (ADEPT) to treat prostate cancer. The kit
 CC comprises several respective containers for holding anti-human seminal
 CC plasma protein single-chain antibody (ScFv)/human carboxypeptidase A
 CC (hCPA) fusion protein (ABP97414), a prodrug of methotrexate-alpha-peptide
 CC and a pharmaceutically acceptable auxiliary for medication. The invention
 CC also encompasses the anti-human seminal plasma protein ScFv/hCPA fusion
 CC protein, and the nucleic acid encoding it (AB275116). This was
 CC constructed by linking DNA encoding the ScFv (itself generated by linking
 CC DNAs encoding the heavy and light chain variable regions (VH and VL) of
 CC the murine anti-human seminal plasma protein monoclonal antibody E4B7)

CC to DNA encoding human carboxypeptidase A via a linker. The ScFv portion
 CC of the fusion protein enables it to be specifically targeted to prostate
 CC cancer cells, where the hCPA portion can then activate the prodrug which
 CC in turn has cytostatic activity towards the cancer cells. The invention
 CC permits the fusion protein to be easily produced in large quantities for
 CC use in ADEPT for prostate cancer treatment. The present sequence
 CC represents the ScFv/hCPA fusion protein of the invention.

XX Sequence 667 AA;

Alignment Scores:

Pred. No.: 1,666-60 Length: 667
 Score: 647,50 Matches: 143
 Percent Similarity: 52.46% Conservative: 81
 Best Local Similarity: 33.49% Mismatches: 132
 Query Match: 23.32% Indels: 72
 DB: 24 Gaps: 10

US-09-980-881A-1 (1-1573) x ABP97414 (1-667)

QY 6 ATTGCTGTTGGGATGAAGCTTTCGACGCTTGCACTGCTGATACCAATGTTCTTCTCTGT 65
 Db 255 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal----- 269
 QY 66 GACGACGATGCTTCGGCTTCCAGAGTGGCCAGTTCTAGCTCTCTCTAGAACCTCT 125
 Db 270 -----ArgPheAspArgGluValPheArgValLysProGlnAspGlu 284
 QY 126 AGCAAGTTCACGATCTTACACAGATCTTACTACACATATGAGATTGTTCTCGCAGCG 185
 Db 285 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheIlePyrPro 304
 QY 186 GTAACAGCTGACCTTATTGGAAGAAACAACTGCTTCTTAAATGTCATCTGAT 245
 Db 305 GlyAlaThrHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 324
 QY 246 GTCGACATGTGAAGCCCATTAATGTGAGCGGAATCCATGCGAGTCTTCTGTCGCA 305
 Db 325 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 344
 QY 306 GACGTGGAGATCTTATCAACAGCAGATTTC---AACGACACAGTCAGCCCCGAGCC 362
 Db 345 AspLeuGlnGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 364
 QY 363 TCCGCATCGTACTGATCAAGTATCACTAAATCAATCTATTCTGGATAGAAATTT 422
 Db 365 Ser-----TyrAlaLysTyrAsnAsnTyrGluLysIleValAlaThrThrGluLys 381
 QY 423 ATAACTGAGAGCATCTGATATGCTTACAAAATCCATTCGATCCTCATTTGAGAAG 482
 Db 382 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 401
 QY 483 TACCCACTCTATGTTTAAAGTTTCTGGAAGAGAACACACGCCAAAATGCCATATGG 542
 Db 402 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgGlyAlaIlePhe 420
 QY 543 ATTGATGTGCAATCCATGCGAGATGATCTCTCTGCTTCTTCTGCTTGTGGTTCATA 602
 Db 421 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 440
 QY 602 ----- 602
 Db 441 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 460
 QY 603 -----GGCCAT-----AAT 611
 Db 461 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 480
 QY 612 CGAATGTGGAAGAACCGCTTCTTCTATGCGACAATCATTCGACATCGGAACAGACCTG 671
 Db 481 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 500
 QY 672 AATAGCAACTTTGCTCCAAACACTGCTGTGAGAGAGGTGCATCCAGTTCTCTCATGCTG 731

Db 501 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 519
 QY 732 GAACCTACTGTGCACTTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTTC 791
 Db 520 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 539
 QY 792 TTGAGAAAGAAATATCAACCAGATTAAAGCATACATCAGCATGATTCATCTCCAGCAT 851
 Db 540 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 559
 QY 852 ATAGTGTTCCTATTTCTTATACACAGCAAGTAAAGCAAGACCATCAGGAACCTGCTCTA 911
 Db 560 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 579
 QY 912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACCTAGTAAATAATACCAAGGTATACACAT 971
 Db 580 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 598
 QY 972 GGCATGGCTCAGAAACCTTATACCTAGCTCTGAGAGTGGGACCGATTGGATCTATGAT 1031
 Db 599 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 618
 QY 1032 TTGGGCATCAATATTCGTTTAC----- 1054
 Db 619 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 638
 QY 1055 -----ATCAAAACCCCTGTAGAGAGCTTTTGCCGCTGTCTCTAAATA 1099
 Db 639 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 658
 QY 1100 GCTTGGCATGTCATTAGGAAT 1120
 Db 659 AlaLysTyrIleLeuLysHis 665

Search completed: January 1, 2004, 18:52:39
 Job time : 231.037 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:40:51 ; Search time 42.3933 Seconds
(without alignments)
3139.880 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 aaaaaattgcttggatg.....aaaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09980881/runat_31122003_135750_24983/app_query.fasta_1.2958
-DB=Issued Patents AA -QFMT=faetan -SUFFIX=rai -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORES=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 @CGN 1 1 31 @runat_31122003_135750_24983 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1902.5	68.5	386	4	US-09-813-133A-2
2	1885	67.9	423	1	US-07-649-591B-3
3	1885	67.9	423	1	US-08-277-540-3
4	1885	67.9	423	1	US-08-430-787A-3
5	1885	67.9	423	2	US-08-869-057-2
6	1874	67.5	423	4	US-09-813-133A-4
7	655	23.6	417	1	US-07-649-591B-7
8	655	23.6	417	1	US-08-277-540-7
9	655	23.6	417	1	US-08-430-787A-7
10	651.5	23.5	415	2	US-08-860-882A-57
11	651.5	23.5	415	4	US-09-011-769A-39
12	647.5	23.3	417	1	US-07-649-591B-6

13	647.5	23.3	417	1	US-08-277-540-6	Sequence 6, Appli
14	647.5	23.3	417	1	US-08-430-787A-6	Sequence 6, Appli
15	644	23.2	424	1	US-09-011-769A-56	Sequence 56, Appl
16	640.5	23.1	404	1	US-08-696-139-2	Sequence 2, Appli
17	635.5	22.9	437	4	US-09-675-305-10	Sequence 10, Appl
18	634.5	22.8	716	3	US-09-171-945-125	Sequence 125, App
19	607.5	21.9	307	2	US-08-782-760-6	Sequence 6, Appli
20	607.5	21.9	307	5	PCI-US96-00995-6	Sequence 6, Appli
21	607	21.9	396	1	US-07-649-591B-4	Sequence 4, Appli
22	607	21.9	396	1	US-08-277-540-4	Sequence 4, Appli
23	607	21.9	396	1	US-08-430-787A-4	Sequence 4, Appli
24	576	20.7	306	1	US-08-696-139-4	Sequence 4, Appli
25	575	20.7	349	4	US-09-011-769A-47	Sequence 47, Appl
26	572.5	20.6	329	4	US-09-011-769A-51	Sequence 51, Appl
27	571.5	20.6	613	3	US-09-171-945-113	Sequence 113, App
28	568	20.5	349	4	US-09-011-769A-60	Sequence 60, Appl
29	567	20.4	349	4	US-09-011-769A-64	Sequence 64, Appl
30	553	19.9	417	1	US-07-649-591B-8	Sequence 8, Appli
31	553	19.9	417	1	US-08-277-540-8	Sequence 8, Appli
32	553	19.9	417	1	US-08-430-787A-8	Sequence 8, Appli
33	546.5	19.7	399	4	US-09-710-099-8	Sequence 12, Appl
34	529.5	19.1	350	4	US-09-675-305-12	Sequence 5, Appli
35	529	19.0	419	1	US-07-649-591B-5	Sequence 5, Appli
36	529	19.0	419	1	US-08-277-540-5	Sequence 5, Appli
37	529	19.0	419	1	US-08-430-787A-5	Sequence 5, Appli
38	520	18.7	436	4	US-09-710-099-6	Sequence 6, Appli
39	513	18.5	417	3	US-08-640-906-4	Sequence 4, Appli
40	513	18.5	417	4	US-09-395-936-4	Sequence 4, Appli
41	512	18.4	417	3	US-08-640-906-18	Sequence 18, Appl
42	512	18.4	417	4	US-09-395-936-18	Sequence 18, Appl
43	487.5	17.6	419	3	US-08-640-906-2	Sequence 2, Appli
44	487.5	17.6	419	4	US-09-395-936-2	Sequence 2, Appli
45	484	17.4	419	3	US-08-640-906-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Alignment Scores:	1.4e-202	Length:	386
Pred. No.:	1902.50	Matches:	363
Score:	94.30%	Conservative:	1
Best Local Similarity:	94.04%	Mismatches:	5
Query Match:	68.51%	Indels:	18
DB:	4	Gaps:	1
US-09-980-881A-1 (1-1573) x US-09-813-133A-2 (1-386)			
QY	18	ATGAAGCTTTGAGCCCTTCAGTCCCTTGATACCATGTTCTCTCTGTGAGCAGCATGTC	77
Db	1	MetLysLeuCySerLeuValValLeuValProlleValLeuPheCysGluGlnHisVal	20
QY	78	TTGCGCTTCCAGAGTGGCCAAAGTTCTAGTGTCTTCTCTAGAACCTCTTAGGCAAGTTC	137

Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 138 GTTCTACAGAACTTACTACACATATGAGATTTCTCTGGCAGCGGTAAACAGCTGAC 197
Db 41 ValLeuGlnAsnLeuThrThrTyrGluValLeuTrpGlnProValThrAlaAsp 60
Qy 198 CTTATTCTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 257
Db 61 LeuIleValIlyslsGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Qy 258 AAAGCCCATTAATGTAGCGGAATTCATGAGTCTCTTGTGGCAGACGTGGAAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 318 CTTATTCAACAGCAGATTTCACAGCACAGTCCAGCCCGAGCCTCCGATCGTACTAT 377
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerTyrTyr 120
Qy 378 GAACAGTATCACTCACTAAATGAATCTATTCTGGATAGAATTTATACTGAGAGCAT 437
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
Qy 438 CTTGATATGCTTCAAAAAATCCCAATTCGATTCCTCATTTGAGAAGTACCCACTATGTT 497
Db 141 ProAspMetLeuThrIlyslsIleHisIleGlySerSerPheGluIlyTyrProLeuTyrVal 160
Qy 498 TTAAGGTTTCTGAAAAGAACACACAGCCCAAAATGCCATATGGATTGATCTGGATC 557
Db 161 LeuIysValSerGlyIysGluGlnAlaAlaIleAsnAlaIleTrpIleAspCysGlyIle 180
Qy 558 CATGCCAGAGAATGGATCTCTCTGCTTCTCTGCTTGTGGTTTCATAGGCCATAATCGAATG 617
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
Qy 618 TGAGAAAGAACCGTCTTCTATCGCAACAATCATTCATCGCAACAGACCTCAATAGC 677
Db 201 TrpArgIysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArg 220
Qy 678 AACTTTGCTCCAAACACTGGTGTGAGAGGTGCATCCAGTTCCTCATCTCGGAAC 737
Db 221 AsnPheAlaSerIlyHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThr 240
Qy 738 TACTGTGACTTTATCTCTAGTTCAGAACCAAGAGTGAAGCGGTAGTCTTCTTGAGA 797
Db 241 TyrCysGlyLeuTyrProGluSerGluProGluValIlyslsAlaValAlaSerPheLeuArg 260
Qy 798 AGAATATCAACAGATTAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTG 857
Db 261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
Qy 858 TTTCATATTCCTATACAGAAAGTAAAGCAAGACCATCAGGAAGTCTCTAGTAGCC 917
Db 281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
Qy 918 AGTGAACAGTTCGTCTATTGACAAACTAGTAAATAATACAGGTATACACATGGCCAT 977
Db 301 SerGluAlaValArgAlaIleGluIlyslsIleSerLysAsnThrArgTyrThrHisGlyHis 320
Qy 978 GGCTCAGAAACCTTATACCTAGCTCTCTGAGAGTGGGACGATTCATGATTTGGGC 1037
Db 321 GlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGly 340
Qy 1038 ATCAATATTCGTT----- 1051
Db 341 IleIlyslsTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360
Qy 1052 -----TACATCAACCCCTGTAGAGAGCTTTTCGCCGTCTCTCTAAATAGCTTGG 1105
Db 361 GluArgTyrIleIlyslsProThrCysArgGluAlaPheAlaValSerLysIleAlaTrp 380
Qy 1106 CATGTCATTAGGAATGTT 1123
Db 381 HisValIleArgAsnVal 386

RESULT 2

US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Alignment Scores:
Pred. No.: 1,29e-200 Length: 423
Score: 1885.00 Matches: 365
Percent Similarity: 86.52% Conservative: 1
Best Local Similarity: 86.29% Mismatches: 3
Query Match: 67.88% Indels: 55
DB: 1 Gaps: 2

US-09-980-881A-1 (1-1573) x US-07-649-591B-3 (1-423)

Qy 18 ATGAAGCTTTCAGACCTTGAGTCTTCAGTCTTGTACCCATTGTTCTCTCTGTGAGCAGCATGTC 77
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Qy 78 TTCGGTTCAGAGTGCCCAAGTTCTAGTGTCTTCTCTAGAACCTCTAGGCAAGTTCAA 137
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 138 GTTCTACAGAACTTACTACACATATGAGATTTCTCTGGCAGCGGTAAACAGCTGAC 197
Db 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
Qy 198 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 257
Db 61 LeuIleValIlyslsGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Qy 258 AAAGCCCATTAATGTAGCGGAATTCATGAGTCTCTTGTGGCAGACGTGGAAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

QY 318 CTTATTCAACAGCAGATTTCACAGCACACAGTCAGCCCCCGAGCTCCGCTACTAT 377
Db 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerTyrTyr 120
QY 378 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGATTATTAACCTGAGAGCAT 437
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 438 CTGTATATGCTTCAAAAATCCCATTTGATTCCTCATTTTCAGAAAGTACCCCACTCTATGTT 497
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 498 TTAAGGTTTCTGGAAAGAACAAACAGCCAAAATGCCATATGCTAGTCTGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 558 CATGCCAGAGAATGGATCTCTCTGCTTCTGCTTGTGGTTTCATAGCCAT----- 608
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 608 ----- 608
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 609 -----AATCGAATGCGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 627 AACGGTTCTTCTATGCGAACATCATTTGCTGCGAACACCTGAATACCACTTTTGTGTC 686
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 687 TCAAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTATGCTCGGAAACCTACTGTGGA 746
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
QY 747 CTTATCTCAGTCAGCAGAACAGAGTGAAGCGAGTGGCTAGTTCTTCTGAGAAAGATATC 806
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 807 AACCAATTAAGCATACATCAGCATGCTTCTATCTACTCCAGCATATAGTGTTCATAT 866
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 867 TCCTATACAGCAAGTAAAGCAAGACCATCAGGAACCTGTCTCTAGTAGCCAGTGAAGCA 926
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAla 340
QY 927 GTTCGTGCTATTGACAAAACCTAGTAAATAACAGGTATACATAGCCCATGGCTCAGAA 986
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 987 ACCTTATACCTAGCTCTGAGGTGGGAGATTGGATCTATGATTTGGGCATCAATAT 1046
Db 361 ThrLeuTyrLeuAlaProGlyGlyAspTrpIleTyrAspTrpIleTyrAspLeuLysTyr 380
QY 1047 TCGTT-----TAC 1054
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1055 ATCAAAACCCACTGTAGAGAGCTTTTGGCGTGTCTCTAAATAGCTTGGCATGTCATT 1114
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLysIleAlaTrpHisValIle 420
QY 1115 AGCAATGTT 1123
Db 421 ArgAsnVal 423

RESULT 3

US-08-277-540-3

; Sequence 3, Application US/08277540

; Patent No. 5474901

; GENERAL INFORMATION:

APPLICANT: Drayna, Dennis T., Baton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-3

Alignment Scores:

Pred. No.:	1,29e-200	Length:	423
Score:	1895.00	Matches:	365
Percent Similarity:	86.52%	Conservative:	1
Best Local Similarity:	86.29%	Mismatches:	3
Query Match:	67.88%	Indels:	55
DB:	1	Gaps:	2

US-09-980-881A-1 (1-1573) x US-08-277-540-3 (1-423)

QY	18	ATCAAGCTTGCAGCTTCAGTCTTGTACCCATTGTTCTTCTGTGTAGCAGCATGTC	77
Db	1	MeuLysLeuCysserLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal	20
QY	78	TTCCGCTTCCAGAGTGCACCAAGTTCTAGCTGCTTCTTAGAACCTCTTAGGCAAGTTCAA	137
Db	21	PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln	40
QY	138	GTTCCTACAGAACTTACTACACATATGAGATTGTTCTCTGGCAGCGCGGTAAACAGCTGAC	197
Db	41	ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaasp	60
QY	198	CTTATTGTGAGAAAAACAAAGTCCATTTTTTTTGTAAATGCATCTCATGTGCAATGTG	257
Db	61	LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValArgAsnVal	80
QY	258	AAAGCCCATTTAAATCTGAGCGCAATTCATGCAGTGTCTTCTGCGCAGACGTGGAGAT	317
Db	81	LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaaspValGluasp	100

Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 318 CTTATTCAACAGCAGATTTCACAGACACAGTCCAGCCCGCCTCCGATCGTACTAT 377
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr 120
Qy 378 GAACAGTATCACATCAATGAATCTATCTGGATAGATTTATTAAGTGAAGCCAT 437
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
Qy 438 CTTGATATGCTTCAAAAAATCCCATTTGATTCCTTCATTTGAGAAGTACCCACTCTATGTT 497
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Qy 498 TTAAGTTTCTGAAAGAACAAACAGCCAAAATGCCATATGATGATCTGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Qy 558 CATGCCAGAGATGGATCTCTCTGCTTCTGCTTCTGCTGGTTCATAGGCCAT----- 608
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
Qy 608 ----- 608
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
Qy 609 -----AATCGAATGTGGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
Qy 627 AACGGTCTTCTATGCGAACATCATTTGATCGGACACACCTGATAGCACTTTGTC 686
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
Qy 687 TCAAAACACTGGTGTGAGGAGGTGCATCCAGTCTCTCATGCTCGGAAACCTACTGTGGA 746
Db 261 SerIysHisTrpCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGly 280
Qy 747 CTTTATCTCGTCAAGTCAAGCAGAGTGAAGCAGTGGCTAGTTCTTTGAGAAGAAATATC 806
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
Qy 807 AACAGATTAAAGCATACATCAGCATGCTATCTATCTCCAGCATATAGTGTTCATAT 866
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
Qy 867 TCCTATACAGAGTAAAGCAAGACCATGAGCACTGCTCTAGTAGCCAGTGAAGCA 926
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
Qy 927 GTTCGTGCTATTGACAAAACCTAGTAAATAATACCAGGTATACACATGGCCATGGCTCAGAA 986
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
Qy 987 ACCTTATACCTAGTCTCGAGGTGGGAGCATTTGATCATATGATTTGGGCATCAATAT 1046
Db 361 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
Qy 1047 TCCTTT-----TAC 1054
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
Qy 1055 ATCAAAACCCACCTGTAGAGAGCTTTTTCGGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLysIleAlaTrpHisValIle 420
Qy 1115 AGGAATGTT 1123
Db 421 ArgAsnVal 423

RESULT 5

US-08-869-057-2

; Sequence 2, Application US/08869057

; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Waentien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-5411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; TISSUE TYPE: Plasma
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 23..401
; US-08-869-057-2

Alignment Scores:
Pred. No.: 1,29e-200 Length: 423
Score: 1885.00 Matches: 365
Percent Similarity: 86.52% Conservative: 1
Best Local Similarity: 86.29% Mismatches: 3
Query Match: 67.88% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-1 (1-1573) x US-08-869-057-2 (1-423)

Qy 18 ATGAAGCTTTCAGCCTTCAGTCCCTTGACCCATTGTTCTCTTGTGAGCAGCATGTC 77
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Qy 78 TTCGCGCTTCAGAGTCCGCAAGTTCTAGTCTCTTCCTAGACCTCTAGGCAAGTTCAA 137
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 138 GTTCTACAGAACTTACTACACATATGAGATTGTTCTCTCGCAGCGCGGTAACAGCTGAC 197
Db 41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
Qy 198 CTTATTGTGAAGAAAAAACAAGTCCATTCTTTTGTAAATGATCTGATGTCGACATGTC 257
Db 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Qy 258 AAGCCCATTTAAATGTGAGCGGAATTCATGACAGTGTCTTGTGGCAGACGTCGAAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

```
QY 318 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCCTCCGATCGTACTAT 377
Db 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
QY 378 GAACAGTATCACTCACTAAATGAATCTTCTTGGATAGAAATTTATACTAGAGGCAT 437
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 438 CTTGATATGCTTCAAAAATCCACATTTGGATCTCTATTTGAGAAGTACCCACTCTATGTT 497
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 498 TTTAAAGTTTCTGGAAGAAGCAACACAGCCAAAATGCCATATGGAATGCTGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 558 CATGCCAGAGAATGGATCTCTCTGCTTCTCTGCTTGTGGTTTCATAGCCCAT----- 608
Db 181 HisAlaArgGlnTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 608 ----- 608
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 609 -----AATCGAATGTCGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 627 AACGGTTCTTTCTATGCGAAACAATCATTTGTCATCGGACACAGACCTGTAATAGCACTTTGTC 686
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 687 TCCAAACACTGGTGTAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
Db 261 SerLysHisTrpCysGluGlyAlaSerSerCysSerSerCysSerGluThrTyrCysGly 280
QY 747 CTTTATCTCAGTCAGAACAGAGTGAAGTGAAGTGCATGCTAGTTCTTGTGAGAAGAAATATC 806
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 300
QY 807 AACGAGATTAAGCATACATCAGCATGCATTCATCTACTCCAGCATATAGTGTTCATAT 866
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 867 TCCTATACAGAGTAAAGCAAGACCATCAGGAATGCTCTAGTAGCCAGTGAAGCA 926
Db 321 SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 340
QY 927 GTTCGTCTATTGCAAAAATAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 987 ACCTTATACCTAGCTCTCGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAAT 1046
Db 361 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1047 TCGTT-----TAC 1054
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1055 ATCAAAACCCACCTAGAGAGCTTTTCCGCTGCTCTCTAAATAGCTTGGCATGTCATT 1114
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLysIleAlaTrpHisValIle 420
QY 1115 AGGAATGTT 1123
Db 421 ArgAsnVal 423
```

RESULT 6

US-09-813-133A-4

; Sequence 4, Application US/09813133A

; Patent No. 6455294

```
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Human
; US-09-813-133A-4

Alignment Scores:
Pred. No.: 2,14e-199 Length: 423
Score: 1874.00 Matches: 363
Percent Similarity: 86.05% Conservative: 1
Best Local Similarity: 85.82% Mismatches: 5
Query Match: 67.48% Indels: 55
DB: 4 Gaps: 2

US-09-980-881A-1 (1-1573) x US-09-813-133A-4 (1-423)
```

```
QY 18 ATGAAGCTTTCAGAGCCTTCAGTCCTTGTACCCATTTGTTCTCTCTGTCAGCAGCATGTC 77
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 78 TTCGCTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTTAGGCAAGTTCAA 137
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 138 GTTCTACAGATCTTACTACACATATGAGATGTTCTCTCGCAGCCGGTAAACAGCTGAC 197
Db 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 198 CTTATTGTCAGAAAAACAAGTCCATTTTGTAAATGCATCTCATGTCGACATGTC 257
Db 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTGGAAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY 318 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCCTCCGATCGTACTAT 377
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
QY 378 GAACGATATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 437
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 438 CTTGATATGCTTCAAAAATCCACATTTGGATCTCTCTATTTGAGAAGTACCCACTCTATGTT 497
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 498 TTTAAAGTTTCTGGAAGAAGCAACACAGCCAAAATGCCATATGGAATGCTGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 558 CATGCCAGAGAATGGATCTCTCTGCTTCTCTGCTTGTGGTTTCATAGCCCAT----- 608
Db 181 HisAlaArgGlnTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 608 ----- 608
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 609 -----AATCGAATGTCGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
```


Qy	627	AACGGTCTTTCTATGCGAACCAATCATTCATTCGGAACAGACCTGGAATAGCAACTTTGTC	686
Db	241	AsnArgSerPheTyrAlaAsnAsnHisCysIleglyThrAspLeuAsnArgAsnPheAla	260
Qy	687	TCCAAACACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGGA	746
Db	261	SerLysHisItrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly	280
Qy	747	CTTTATCTCAGTCAGAACCAACGAAGTCAAGCAGCTGGCTAGTTTCTTTGAGAAGAAATATC	806
Db	281	LeuTyrProGluSerGluProGluValIyAlaValAlaSerPheLeuArgArgAsnIle	300
Qy	807	AACCAGATTAAAGCATACATCAGCATGCATTCACTCCAGCATATAGTGTTCCTCATAT	866
Db	301	AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr	320
Qy	867	TCCTATACAGAGTAAAGCAAGACCATGAGGAAGTGTCTCTAGTAGCCAGTGAAGCA	926
Db	321	SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla	340
Qy	927	GTTTCGTGCTATTGCACAAACTAGTATAAAATACCAAGGTATACACATGGCCATGGCTCAGAA	986
Db	341	ValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu	360
Qy	987	ACCTTATACCTAGTCTCTCGAGGTGGGAGCATTTGGATCTATGATTTGGCATCAAAATAT	1046
Db	361	ThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGlyIleLysTyr	380
Qy	1047	TCGTTT-----TAC	1054
Db	381	SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr	400
Qy	1055	ATCAAAACCCACCTGTAGAGAAGCTTTTGGCCGCTGCTCTAAATATAGCTTGGCATGTCAATT	1114
Db	401	IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle	420
Qy	1115	AGGAATGTT	1123
Db	421	ArgAsnVal	423

RESIN.T 7

```

US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haeak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:

```

```
Db 259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278
      ::::: ||| ||| ||| |||
Qy 759 TCAGACAGAGTGAAGGAGTGGCTAGTCTTCTTGAGAGAAATATCAACGAGTAAA 818
      ::::: ||| ||| ||| |||
Db 279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
      ::::: ||| ||| ||| |||
Qy 819 GCATACATCAGCATGCTATCTATCCAGCATATAGTGTTCATATTCCTATACAGA 878
      ::::: ||| ||| ||| |||
Db 299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318
      ::::: ||| ||| ||| |||
Qy 879 AGTAAAGCAAGACCATGAGCACTCTCTAGTAGCCAGTGAAGCAGTTCGTCTATT 938
      ::::: ||| ||| ||| |||
Db 319 LysLeuProProAsnHisGlnAspLeuLysValAlaArgIleAlaThrAspAlaLeu 338
      ::::: ||| ||| ||| |||
Qy 939 GACAAACTAGTAAATACAGTATACATGAGCCATGAGCCATGCTACAGAACTTATACCTA 998
      ::::: ||| ||| ||| |||
Db 339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357
      ::::: ||| ||| ||| |||
Qy 999 GCTCTGGAGGTGGGAGCAGTATGATCTATGATTGGGCATCAAAATTCGTTTAC--- 1054
      ::::: ||| ||| ||| |||
Db 358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377
      ::::: ||| ||| ||| |||
Qy 1055 -----ATCAAAACCCACC 1066
      ::::: ||| ||| ||| |||
Db 378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397
      ::::: ||| ||| ||| |||
Qy 1067 TGAGAGAACTTTGGCGCTGCTCTAAATAGTCTGGCATGCTATTAGGAAT 1120
      ::::: ||| ||| ||| |||
Db 398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415
      ::::: ||| ||| ||| |||

RESULT 8
US-08-277-540-7
; Sequence 7, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasek, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-277-540-7

Alignment Scores:
Pred. No.: 6,12e-64 Length: 417
Score: 655.00 Matches: 150
Percent Similarity: 51.67% Conservative: 66
Best Local Similarity: 35.89% Mismatches: 138
Query Match: 23.59% Indels: 65
DB: 1 Gaps: 9

US-09-980-881A-1 (1-1573) x US-08-277-540-7 (1-417)
Qy 33 CTTCAGAGTCTCTGTACCCATTGTTCTTCTGTGAGCAGCATGTTCTCGGTTCCAGAGT 92
      ::::: ||| ||| ||| |||
Db 7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis-----PheAspArg 23
      ::::: ||| ||| ||| |||
Qy 93 GGCCAAAGTTCTAGCTGCTCTCTCTAGAACCTCTAGGCAAGTTCAAGTCTTACAGATCTT 152
      ::::: ||| ||| ||| |||
Db 24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43
      ::::: ||| ||| ||| |||
Qy 153 ACTACAAACATATGAGATTGTTCTCTGGCAGCCGTAACAGCTGACCTTATTGTGAAGAA 212
      ::::: ||| ||| ||| |||
Db 44 ThrGlnSerIleGluLeuAspPheTyrProAspAlaIleHisAspIleAlaValAsn 63
      ::::: ||| ||| ||| |||
Qy 213 AAACAAGTCCATTTTGTAAATGATCTGATGTCGACAATGTGAAGCCCACTTTAAAT 272
      ::::: ||| ||| ||| |||
Db 64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83
      ::::: ||| ||| ||| |||
Qy 273 GTGAGCGGAATCCATGAGTGTCTGTCGGCAGAGCTGGAAGATCTTATTCAACAGCAG 332
      ::::: ||| ||| ||| |||
Db 84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluIleGluLysGln 103
      ::::: ||| ||| ||| |||
Qy 333 ATTTCC---AACGACACAGTCAGCCCGCCGCTCGCATCGTACTATGAACAGATACAC 389
      ::::: ||| ||| ||| |||
Db 104 PheAspValLysAspGluIleAlaGlyArgHisSer-----TyrAlaLysTyrAsn 120
      ::::: ||| ||| ||| |||
Qy 390 TCATAAATGAATCTATTCTTGATAGAAATTTTAACTGAGAGGATCTCTGATATGCTT 449
      ::::: ||| ||| ||| |||
Db 121 AspTyrAspLysIleValSerTyrThrGluLysMetLeuGluLysHisProGluMetVal 140
      ::::: ||| ||| ||| |||
Qy 450 ACAAATCCACATTCGATCCTCATTTGAGAAGTACCCACTCTATCTTTTAAAGTTTCT 509
      ::::: ||| ||| ||| |||
Db 141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159
      ::::: ||| ||| ||| |||
Qy 510 GGAAGAGAACAAACAGCCAAATGCATATGATGATTGACTGTGGAATCCATGCGCAGAA 569
      ::::: ||| ||| ||| |||
Db 160 GlyLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179
      ::::: ||| ||| ||| |||
Qy 570 TGGATCTCTCTGCTTTCTGTTGTGTTTCATA-----GGCCAT 608
      ::::: ||| ||| ||| |||
Db 180 TrpIleSerProAlaPheCysGlnTyrPheValTyrGlnAlaThrLysSerTyrGlyLys 199
      ::::: ||| ||| ||| |||
Qy 609 AAT----- 611
      ::::: ||| ||| ||| |||
Db 200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219
      ::::: ||| ||| ||| |||
Qy 612 -----CGAATGTGGAGAAAGAACCGTCTTTTC 638
      ::::: ||| ||| ||| |||
Db 220 ValAspGlyTyrIleTyrSerTyrThrGlnAspArgMetTyrArgLysAsnArgSerArg 239
      ::::: ||| ||| ||| |||
Qy 639 TATGCGAAACAATCATTTGATCGGAACAGACCTGAATAGCAACTTTCTCTCCAACACTGG 698
      ::::: ||| ||| ||| |||
Db 240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTyr 258
      ::::: ||| ||| ||| |||
Qy 699 TGTGAGAGAGGTGCATCCAGTCTCTCATGCTCGGAAACCTACTGTGAGACTTTATCTCTGAG 758
      ::::: ||| ||| ||| |||
Db 259 AspSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278
      ::::: ||| ||| ||| |||
Qy 759 TCAGAACCAAGAGTGAAGGAGTGGCTAGTTTCTTCGAGAGAAATATATCAACGAGTAAA 818
      ::::: ||| ||| ||| |||
```

```
Db 279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAenSerIleLys 298
Qy 819 GCATACATCAGTCATTCATCCAGCATATAGTGTTCATATCTTATACACGA 878
Db 299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuProTyrGlyTyrThrPhe 318
Qy 879 AGTAAAGCAAGACCATAGGAGTCTCTCTAGTACGACAGTGAAGAGTTCGTATT 938
Db 319 LysLeuProProHisGlnAspLeuLysValAlaArgIleAlaThrAspAlaLeu 338
Qy 939 GACAAACTAGTAAATACCATACATACATGCGCCATCGCTCAGAAACCTTATACCTA 998
Db 339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357
Qy 999 GCTCTGGAGGTGGGACGATCTGATCTGATGTTGGCATCAATATTCGTTTAC --- 1054
Db 358 ThrSerGlySerSerLeuAspTyrValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377
Qy 1055 -----ATCAAAACCCACC 1066
Db 378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397
Qy 1067 TGTAGAGAGCTTTTGGCGCTGCTCTCTAAATAGCTTGGCATCTCATTAGGAAT 1120
Db 398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415

RESULT 9
US-08-430-787A-7
; Sequence 7, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430, 787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-430-787A-7

Alignment Scores:
Pred. No.: 6,12e-64 Length: 417
Score: 655.00 Matches: 150
Percent Similarity: 51.67% Conservative: 66
Best Local Similarity: 35.89% Mismatches: 138
Query Match: 23.59% Indels: 65
DB: 1 Gaps: 9

US-09-980-881A-1 (1-1573) x US-08-430-787A-7 (1-417)
Qy 33 CTTGAGTCTCTTACCCATTGTTCTTCTGTGAGCAGCATGTCTTCGGTTCACAGT 92
Db 7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis -----PheAspArg 23
Qy 93 GGCCAAAGTCTAGCTCTCTCTAGAACCTCTAGGCAAGTTCAAGTCTTACAGATCTT 152
Db 24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43
Qy 153 ACTACAAACATATATGATTTGTTCTCTGCGACCGGTAACAGCTGACCTTATTGTGAAGAA 212
Db 44 ThrGlnSerIleGluLeuAspPheTyrProAspAlaIleHisAspIleAlaValAsn 63
Qy 213 AAACAAGTCCATTTTGTAAATGTCATCTGATGTCGACAAATGTGAAGCCCATTTAAAT 272
Db 64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83
Qy 273 GTGAGCGGAATTCATGTCAGTGTCTGTGCGACAGCTGGAAGATCTTATCAACAGCAG 332
Db 84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluIleGluLysGln 103
Qy 333 ATTTCC---AACGACACAGTCAGCCCCCGACCTCGCATCTGCTACTATGAACAGTATCAC 389
Db 104 PheAspValLysAspGluIleAlaGlyArgHisSer-----TyrAlaLysTyrAsn 120
Qy 390 TCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGGATCTCTGATATGCTT 449
Db 121 AspTyrAspLysIleValSerTyrThrGluLysMetLeuGluLysHisProGluMetVal 140
Qy 450 ACAAATCCATTCATGTCATTTGAGAAATGACCTGCTCTATGTTTAAAGTTTCT 509
Db 141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle-- 159
Qy 510 GGAAGAAGAACACACAGCCAAATGCAATGATGCTGCTGGAATCCATGCCAGAGAA 569
Db 160 GlyLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179
Qy 570 TGGATCTCTCTCTTCTGTTGTTGTTTCATA-----GGCCAT 608
Db 180 TrpIleSerProAlaPheCysGlnTyrPheValTyrGlnAlaThrLysSerTyrGlyLys 199
Qy 609 AAT----- 611
Db 200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219
Qy 612 -----CGAATCTGGAGAAAGACCTTCTTTC 638
Db 220 ValAspGlyTyrIleTyrSerTyrThrGlnAspArgMetTyrArgLysAsnArgSerArg 239
Qy 639 TATGCCAAACAATCATTTGTCATCGGAACAGACCTGAATAGCAACTTTGTCTCAAAACACTGG 698
Db 240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTyr 258
Qy 699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTTACTGTGGACTTTATCTCGAG 758
Db 259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278
Qy 759 TCAGAACCCAGAGTGAAGCGAGTGGCTAGTTCTTCTGAGAAGAATAATCAACAGATATAA 818
Db 279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
```

819 GCATACATCAGCATGCTCCAGCATATAGTGTTCATATTCCTATACAGA 878
|||||
299 AlatyrllethPheHisSerTySerGlnMetLeuLeuLeuLeuLeuLeuPhe 318
|||||
879 AGTAAAGCAAGACCATGAGCACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
|||||
319 LysLeuProProAenHisGlnAspLeuLeuLysValAlaAlaArgIleAlaThrAspAlaLeu 338
|||||
939 GACAAACTAGTAAATAATACAGGTATACATGCGCCATGCTCAGAAACCTTATACCTTA 998
|||||
339 ---SerThrArgTyrgluThrArgTyrlleTyrglyProIleAlaSerThrIleTyrls 357
|||||
999 GCTCTCGGTGGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1054
|||||
358 ThrSerGlySerSerLeuAspTrpValTyAspLeuGlyIleLysHisThrPheAlaPhe 377
|||||
1055 -----ATCAACCCACC 1066
|||||
378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProgluSerArgIleLysProThr 397
|||||
1067 TGTAGAGAGCTTTGCGGCTGCTCTTAAATAGCTTGGCATGCTATTAGGAAT 1120
|||||
398 CysLysgluThrMetLeuSerValLysPheIleAlaLysTyrlleLeuLysAsn 415
|||||

RESULT 10
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HOW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-860-882A-57

Alignment Scores: 1.5e-63 Length: 415
Score: 651.50 Matches: 142
Percent Similarity: 52.03% Conservative: 76
Best Local Similarity: 33.89% Mismatches: 140
Query Match: 23.46% Indels: 62
DB: 2 Gaps: 8

US-09-980-881A-1 (1-1573) x US-08-860-882A-57 (1-415)

QY 33 CTTCAGTCTCTTGTACCCATTTGCTCTCTGTGAGCAGCATCTCTTCGCG---TTTCAG 89
|||
Db 2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisGlyGlyGluHisPheGlu 21
|||
QY 90 AGTGGCAAGTCTTACGCTGCTCTCTCTAGAACCTCTAGCACTTCAAGTCTTACAGAT 149
|||
Db 22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleArgGlu 41
|||
QY 150 CTTCATACACATATGAGATTTGCTCTGCGACGCGTAACAGCTGACCTTATTGTGAAG 209
|||
Db 42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61
|||
QY 210 AAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCACAAATGTGAAAGCCCATTTA 269
|||
Db 62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81
|||
QY 270 AATGTGCGGGAATTCATGCGATGCTGTCTGCGCAGAGCTGGAAGATCTTATTCAACAG 329
|||
Db 82 LysGlnAsnGluLeuGlnTyrlsValLeuIleSerAsnLeuArgAsnValValGluAla 101
|||
QY 330 CAGATTTCCACACACACAGTCAGCCCGAGCTCCGCATCGTACTATGAACAGTATCAC 389
|||
Db 102 GlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrgluLysTyrlsAsn 118
|||
QY 390 TCACTAAATGAATCTATTCTGGATAGATTTTAACTGAGAGGATCTCTGATATGCTT 449
|||
Db 119 LysTrpGluThrIleGluAlaTrpThrGlnValAlaThrGluAsnProAlaLeuIle 138
|||
QY 450 AAAAAATCCACATTCGATCCTCATTTGAGAGTACCCACTCTATGTTTAAAGTTTCT 509
|||
Db 139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrlsLeuLysVal--- 157
|||
QY 510 GGNAAAGAACAAACAGCCCAATGCAATGATGATGATGATGATGATGATGATGATGATGAT 569
|||
Db 158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177
|||
QY 570 TGGATCTCTCTGCTTTTCTGCTTGTGTTTCATA----- 602
|||
Db 178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrglyArg 197
|||
QY 602 -----GGCCAT----- 602
|||
Db 198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrlsValLeuProValLeuAsn 217
|||
QY 603 -----GGCCAT-----AATCGAATGTGGAGAAAGAACCTTCTTTC 638
|||
Db 218 IleAspGlyTyrlleTyrlsThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237
|||
QY 639 TATGCGAACAAATCATTTGCAATCGGAACAGACCTGCAATAGCAACTTGTCTCAACACTGG 698
|||
Db 238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256
|||
QY 699 TGTGAGGAAGGTGCATCCAGTCTCTCATGCTCGGAAACCTTACTGTGGACTTTTCTCTGAG 758
|||
Db 257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrlsGlyProAlaAlaGlu 276
|||
QY 759 TCAGAACCAAGTGAAGCAGTGGCTAGTCTTCTGAGAGAGAAATATCAACAGATTA 818
|||
Db 277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296
|||
QY 819 GCATACATCAGCATGCTATTCATCTCCAGCATATAGTGTTCATATTCCTATACAGA 878
|||

Db 297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316
Qy 879 AGTAAAGCAAGACCATGAGCACTGTCTCTAGTAGCCAGTGAAGCAGTGTCTGTCTATT 938
Db 317 LysLeuGlyGluAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336
Qy 939 GACAAACTAGTAAATACCATGATATACATGGCCATGGCTCAGAAACCTTATACCTA 998
Db 337 ---AlaSerLeuHisGlyThrLysTyrThrGlyProGlyAlaThrThrIleTyrPro 355
Qy 999 GCTCTCGAGGTGGGACGATCTGATGATTGGGCATCAATATCTGTTTACATC- 1057
Db 356 AlaAlaGlyGlySerAspThrAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375
Qy 1058 -----AAACCCACC 1066
Db 376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGlnSerGlnIleArgAlaThr 395
Qy 1067 TGTAGAGAGCTTTTGGCGTGTCTCTAAATAGCTTGGCATCTCATTAGGAATGTT 1123
Db 396 CysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHisLeu 414

RESULT 11
US-09-011-769A-39
; Sequence 39, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAW, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011.769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-011-769A-39

Alignment Scores:
Pred. No.: 1.5e-63 Length: 415

Score: 651.50 Matches: 142
Percent Similarity: 52.03% Conservative: 76
Best Local Similarity: 33.89% Mismatches: 140
Query Match: 23.46% Indels: 62
DB: 4 Gaps: 8

US-09-980-881A-1 (1-1573) x US-09-011-769A-39 (1-415)

Qy 33 CTTGCACTCTTTGTACCAATTTGTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 89
Db 2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisGlyGlyGluHisPheGlu 21
Qy 90 AGTGGCAAGTTCTAGCTCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAT 149
Db 22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleLeuArg 41
Qy 150 CTTTACTACAACTATGAGATTGTTCTCTGCGAGCCGTAACAGCTGACCTTATTGTGAAG 209
Db 42 LeuAlaSerThrThrGlnIleAspPheThrLysProAspSerValThrGlnIleLysPro 61
Qy 210 AAAAAACAAGTCCATTTTGTAAATGCTCTGATGTCGACAAATGTGAAGCCCATTTA 269
Db 62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81
Qy 270 AATGTGAGCGGAATTCATGTCAGTGTCTGTCGAGAGCTGGAAGATCTTATCAACAG 329
Db 82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101
Qy 330 CAGATTTTCAACGACACAGTCAGCCCCCGAGCTCCGCATCGTACTATGACAGATGATC 389
Db 102 GlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118
Qy 390 TCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGGATCTCTGATGCTT 449
Db 119 LysTrpGluThrIleGluAlaThrThrGlnValAlaThrGluAsnProAlaLeuIle 138
Qy 450 ACAAAATCCCATGTCATGTCATTTGAGAGTACCCACTCTATGTTTAAAGTTTCT 509
Db 139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal 157
Qy 510 GGAAGAACAACAGCCCAAAATGCCATATGATGTCATGTCGATCCATCCAGAGAA 569
Db 158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArg 177
Qy 570 TGGATCTCTCTGCTTTCTGCTTGTGCTTCATA----- 602
Db 178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197
Qy 602 ----- 602
Db 198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217
Qy 603 -----GGCCAT-----AATCGAATGTGGAAGAAGACCTTCTTTC 638
Db 218 IleAspGlyTyrIleTyrThrTrpLysSerArgPheThrArgLysThrArgSerThr 237
Qy 639 TATGCAACAATCATTGTCATCGAAGACCATGTAATGAGCACTTCTCTCAAACTGG 698
Db 238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256
Qy 699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCGAG 758
Db 257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276
Qy 759 TCAGAACCAAGAGTGAAGCAGTGGCTAGTTTCTTGAGAAGAATATCAACAGATTAA 818
Db 277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296
Qy 819 GCATACATCAGATGCTATCCAGCATATAGTGTTCCTCATATTCTTATCCATACACA 878
Db 297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316
Qy 879 AGTAAAGCAAGACCATGAGCACTGTCTCTAGTAGCCAGTGAAGCAGTGTCTGTCTATT 938

```
Db 317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 939 GACAAACTAGTAAATACAGGTATACATGGCCATGGCTCAGAACTTATACCTA 998
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 337 ----AlaSerLeuHisGlyThrLysThrTyrGlyProGlyAlaThrThrIleTyrPro 355
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 999 GCTCTGGAGTGGGACGATTGATCTATGATTGGGCATCAAAATTCGTTTACATC- 1057
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 356 AlaAlaGlySerAspPTrPalatYrAspGlnGlyIleArgTyrSerPheThrPhe 375
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 1058 -----AAACCCACC 1066
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 1067 TGTAGAGAACTTTGCGCTGCTCTAAATAGCTGGCATGTCTATTAGGAATGT 1123
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 396 CysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHisLeu 414
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
RESULT 12
US-07-649-591B-6
; Sequence 6, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haseak, Janet B.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
US-07-649-591B-6
Alignment Scores:
Pred. No.: 4,17e-63 Length: 417
Score: 647.50 Matches: 143
Percent Similarity: 52.46% Conservative: 81
Best Local Similarity: 33.49% Mismatches: 132
Query Match: 23.32% Indels: 72
DB: 1 Gaps: 10
US-09-980-881A-1 (1-1573) x US-07-649-591B-6 (1-417)
QY 6 ATTGCTGTTGGGATGAAGCTTTTGACGCTTTCGAGTCTCTGTACCCATGTTCCTCTGT 65
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
```

```
Db 5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle----AlaProVal----- 19
QY 66 GAGCAGCATGTCTTCCGGTTCAGAGTGGCCAGATTCTAGCTGCTCTTCTCAGAACCTCT 125
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 20 -----ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 126 AGCAAGTTTCAAGTTCTACAGAACTTCTACTACCAATATGAGATTGTCTCTCGCAGCCG 185
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheThrTyrPro 54
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 186 GTAAACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGTCATCTGAT 245
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 246 GTCGCAATGTGNAAGCCCATTTAAATCTGACGGGAATTCATGCGAGTGTCTTGTGGCA 305
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 306 GAGCTGGAAGATCTTATTCAACAGCAGATTTC-----AACGACACAGTCCAGCCCCGAGCC 362
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 363 TCCGCATCGTACTATCAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTT 422
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 115 Ser-----TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaIleThrThrGluLys 131
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 423 ATAACTGAGAGCATCTCTGATATGCTTACAAAATCCACATTCGATCTCTATTGTAGAAG 482
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 483 TACCCACTCTATGTTTTAAAGTTTCTGGAAGAAACAAACAGCCAAAATGTCATATGG 542
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 152 AsnProLeuTyrValLeuLysIle---GlyLysAsnGluArgLysAlaIlePhe 170
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 543 ATTGACTGTGAATCCATCCAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTA 602
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 602 ----- 602
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 603 -----GGCAT-----AAT 611
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 612 CGAATGTGGAGAAAGAACCGTTCTTCTATCGAACAAATCATTCATCGAACAGACAGCTG 671
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 672 AATAGCAACTTTGTCTCCAAACACTCGTGTGAGGAAGGTGCATCCAGTTCCTCATCTCG 731
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 732 GAAACCTACTGTGGCTTTTATCTGAGTCAGAACAGAGTAGTGAAGCAGTGCGTAGTTTC 791
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 792 TTGAGAAGAAATATCAACAGATTAAGCATATACATCAGCATGTCATTCATCTCCAGCAT 851
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 852 ATAGTGTTCATATTCTCTATACCAAGTAAAGCAAGACCATCAGGACTGTCTCTA 911
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 912 GTAGCCAGTGAACAGTTCGTGCTATTGACAAAACCTAGTAAAAAATACCAGGTATACAT 971
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
QY 972 GGCCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACGATGTGATCTATGAT 1031
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
Db 349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368
      :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
```


Db 349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368
Qy 1032 TTGGGCATCAATATTCGTTTAC----- 1054
Db 369 LeuGlyIleLeuHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 398
Qy 1055 -----ATCAACCCACCTGTAGAGAACCTTTTGGCGGTCTCTCAAAATA 1099
Db 389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408
Qy 1100 GCTTGGCATGTCATTAGGAAT 1120
Db 409 AlaLysTyrIleLeuLysHis 415

RESULT 15

US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter I.
; DOWELL, Robert R.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011.769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-011-769A-56

Alignment Scores:
Pred. No.: 1.03e-62 Length: 424
Score: 644.00 Matches: 136
Percent Similarity: 52.62% Conservative: 75
Best Local Similarity: 33.92% Mismatches: 130
Query Match: 23.19% Indels: 61
DB: 4 Gaps: 7

US-09-980-881A-1 (1-1573) x US-09-011-769A-56 (1-424)

Qy 84 TTCAGAGTGGCCAGTTCTAGCTGCTCTTCTAGAACCTCTTAGGCAAGTTCAAGTTCTA 143
Db 29 PheGluGlyLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIle 48
Qy 144 CAGAATCTTACTACACATATGAGATTGTTCTGCGCAGCCGGTAACAGCTGACCTTATT 203
Db 49 ArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIle 68
Qy 204 GTGAAGAAAAACAAGTCCATTTTTTTGTAATGTCATCTCATGTCGACAAATGTGAAGCC 263
Db 69 LysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsn 88
Qy 264 CATTTAAATGTGAGCGAATTCCATGCAAGTGTCTGCTGCGCAGACGTGGAAGATCTTATT 323
Db 89 ValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValVal 108
Qy 324 CAACAGCAGATTTTCCACAGACACAGTCAGCCCGCAGCCTCCGCATCGTACTATGAACAG 383
Db 109 GluAlaGlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGluLys 125
Qy 384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTTATAACTGAGAGGCATCTGTAT 443
Db 126 TyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAla 145
Qy 444 ATGCTTACAAAAATCCACATTTGATGCTCTCAATTTGAGAAGTACCACCTCTATGTTTAAAG 503
Db 146 LeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLys 165
Qy 504 GTTCTCTGGAAGAACAAACAGCCAAATGCCATATGGATTGACCTGGAATCCATGCC 563
Db 166 Val---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAla 184
Qy 564 AGAGAATGATCTCTCTGCTTTCTGCTTGTGTTGTTTCA----- 602
Db 185 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyr 204
Qy 602 ----- 602
Db 205 GlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProVal 224
Qy 603 -----CGCCAT-----AATCGAATGTGGAGAAAGACCGT 632
Db 225 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArg 244
Qy 633 TCTTTCTATCGGAACAATCATTCATTCGGAACAGCTGAATAGCACTTTGTCTCCAAA 692
Db 245 SerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAla 263
Qy 693 CACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAACCTACTCTGGACTTTAT 752
Db 264 GlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAla 283
Qy 753 CCTGAGTCAAGACCAAGTGAAGGAGTGGCTAGTTCTTGTGAGAGAATAATCAACCAG 812
Db 284 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSer 303
Qy 813 ATTAAGACATACATCAGCATTCATCTCCAGCATATAGTGTTCATATTCCTAT 872
Db 304 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyr 323
Qy 873 ACACGAAGTAAAGCAAGACCATGAGAACCTGCTCTAGTAGCCAGTGAAGCAGTTCTGT 932
Db 324 AlaTyrLysLeuGlyLysAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLys 343
Qy 933 GCTATTGACAAAACCTAGTAAATACACAGTATACATCGCCATGGCCATGGCTCAGAACTTA 992
Db 344 GluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIle 362
Qy 993 TACCTAGTCTCTGGAGTGGGACGATTCGATCTATGATTGGGCATCAATATTCGTTT 1052
Db 363 TyrProAlaAlaGlyLysSerAspTrpAlaTyrAspGlnGlyIleArgTyrSerPhe 382
Qy 1053 ACATC-----AAA 1060

```

Db      383 ThrPheGluLeuArgAspThrGlyArgTyGlyPheLeuLeuProGluSerGlnIleArg 402
      |||
QY      1061 CCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAATAAGCTTGGCATGTCATTAGGAAT 1120
      |||||
Db      403 AlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHis 422
      |||||
QY      1121 GTT 1123
      |||
Db      423 Leu 423

```

Search completed: January 1, 2004, 19:03:08
 Job time : 69.3933 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:59:01 ; Search time 69.6674 Seconds
(without alignments)
9060.445 Million cell updates/sec

Title: US-09-980-881a-1

Perfect score: 2777

Sequence: 1 agaaaattgcttggatg.....aaaaaaaaaaaaaaaaaaaaa 1573

Scoring table:

BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 73937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 1467874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US0980881/runat_31122003_135753_25135/app.query.fasta_1.2958
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAFSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO MMAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSPBLOCK=100_25135
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

1	1885	67.9	423	12	US-10-379-836-17	Sequence 17, Appl
2	1796	64.7	423	12	US-10-379-836-2	Sequence 2, Appl
3	1592.5	57.3	422	12	US-10-379-836-18	Sequence 18, Appl
4	1568.5	56.5	422	12	US-10-379-836-16	Sequence 16, Appl
5	889	32.0	211	9	US-09-925-302-467	Sequence 467, App
6	647.5	23.3	417	12	US-10-341-434-188	Sequence 188, App
7	647.5	23.3	417	15	US-10-229-546-2	Sequence 2, Appl
8	647.5	23.3	417	15	US-10-229-546-9	Sequence 9, Appl
9	644	22.9	402	12	US-10-379-836-20	Sequence 20, Appl
10	635.5	22.9	437	12	US-10-274-639-12	Sequence 12, Appl
11	635.5	22.9	437	14	US-10-200-344-10	Sequence 10, Appl
12	634.5	22.8	716	10	US-09-910-059-125	Sequence 125, App
13	571.5	20.6	613	10	US-09-910-059-113	Sequence 113, App
14	570	20.3	374	9	US-09-888-615-61	Sequence 61, Appl
15	563	20.3	444	16	US-10-176-306-74	Sequence 74, Appl
16	546.5	19.7	399	15	US-10-200-910-8	Sequence 8, Appl
17	529.5	19.1	350	14	US-10-200-344-12	Sequence 12, Appl
18	520	18.7	436	15	US-10-200-910-6	Sequence 6, Appl
19	512	18.4	436	14	US-10-076-535-2	Sequence 2, Appl
20	508	18.3	417	9	US-09-923-779-150	Sequence 150, App
21	504	18.1	428	9	US-09-925-297-528	Sequence 528, App
22	497	17.9	298	12	US-10-379-836-19	Sequence 19, Appl
23	489.5	17.6	406	9	US-09-925-297-517	Sequence 517, App
24	489.5	17.6	419	12	US-10-345-680-65	Sequence 65, Appl
25	483	17.4	421	11	US-09-948-374-234	Sequence 234, App
26	483	17.4	421	12	US-10-015-387A-234	Sequence 234, App
27	483	17.4	421	12	US-10-006-130A-234	Sequence 234, App
28	483	17.4	421	12	US-10-199-672-308	Sequence 308, App
29	483	17.4	421	12	US-10-006-172A-234	Sequence 234, App
30	483	17.4	421	12	US-10-187-749-308	Sequence 308, App
31	483	17.4	421	12	US-10-194-457-308	Sequence 308, App
32	483	17.4	421	12	US-10-184-642-308	Sequence 308, App
33	483	17.4	421	12	US-10-196-747-308	Sequence 308, App
34	483	17.4	421	12	US-10-015-392A-234	Sequence 234, App
35	483	17.4	421	12	US-10-017-253A-234	Sequence 234, App
36	483	17.4	421	12	US-10-173-689-308	Sequence 308, App
37	483	17.4	421	12	US-10-173-690-308	Sequence 308, App
38	483	17.4	421	12	US-10-173-691-308	Sequence 308, App
39	483	17.4	421	12	US-10-173-692-308	Sequence 308, App
40	483	17.4	421	12	US-10-173-694-308	Sequence 308, App
41	483	17.4	421	12	US-10-173-698-308	Sequence 308, App
42	483	17.4	421	12	US-10-173-699-308	Sequence 308, App
43	483	17.4	421	12	US-10-173-707-308	Sequence 308, App
44	483	17.4	421	12	US-10-174-569-308	Sequence 308, App
45	483	17.4	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1

US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAPI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Alignment Scores:	5.6e-179	Length:	423
Pred. No.:	1885.00	Matches:	365

Percent Similarity: 86.52% Conservative: 1
Best Local Similarity: 86.29% Mismatches: 3
Query Match: 67.88% Indels: 55
DB: 12 Gaps: 2

US-09-980-881A-1 (1-1573) x US-10-379-836-17 (1-423)

QY 18 ATGAAGCTTTGACGCTTGCAGTCTCTGTATCCCATTTCTCTCTCTGAGCAGCATGTC 77
DB 1 MetLysLeuCysSerLeuAlaValLeuValProLeuValLeuPheCysGluGlnHisVal 20

QY 78 TTCGCGTTCAGAGTGCACCAAGTTCTAGCTCTCTCTAGAACCTCTAGGCAAGTTCAA 137
DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40

QY 138 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGCGAGCCGGTAACAGCTGAC 197
DB 41 ValLeuGlnAsnLeuThrThrTyrGluValLeuTrpGlnProValThrAlaAsp 60

QY 198 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 257
DB 61 LeuLeuValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

QY 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAATGTCGACAAATGTG 317
DB 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

QY 318 CTTATTCAACGACGATTTCCACGACACAGTACGCTCTGCTGAGAGTCTCTGCTGAGCAGCTGGAAGAT 377
DB 101 LeuLeuGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

QY 378 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAATTTATACTGAGAGCAT 437
DB 121 GluGlnIleHisSerLeuAsnGluIleTyrSerTrpLeuGluPheIleThrGluArgHis 140

QY 438 CCTGATATGCTTCAAAAAATCCACATTCGATTCCTCTTGGAGAGTACCCCACTCTATGTT 497
DB 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

QY 498 TTAAGGTTTCTGAAAAAACAACAACAGCCAAAATGCCATATGATGATGCTGGGAATC 557
DB 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

QY 558 CATGCCAGAGATCGATCTCTCGCTTCTGCTGTTGGTTTCATAGGCCAT----- 608
DB 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

QY 608 ----- 608

DB 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

QY 609 -----AATCGAATGTGAGAAAG 626

DB 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

QY 627 AACGTTCTTTCTATGCGAACAACTATTGATCGGACACAGACCTGATAGCACTTTGTC 686

DB 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

QY 687 TCCAAACACATCGGTGTGAGGAGGTGCATCCAGTTCTCATGCTCGGAAACCTACTGTGGA 746

DB 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280

QY 747 CTTTATCTGAGTGCAGAACCGAAGTGAAGCAGTGGCTAGTTCTTCTGAGAGAAATATC 806

DB 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

QY 807 AACGAGATTAAGCATACATCAGCATGCTTCACTCCAGCATATAGTTTCCATAT 866

DB 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

QY 867 TCCTATACAGAAAGTAAAGCAAGACCATGAGGAAGTCTCTTAGTAGCCAGTGAAGCA 926

DB 321 SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 340

QY 927 GTTCGCTGCTATTGACAAACTAGTAAATACACAGGTATACACATGCGCATGCTCAGAA 986

DB 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

QY 987 ACCTTATACCTAGCTCTCTGAGGTGGGACGATTGGATCTATGATTGGGCATCAAAATAT 1046

DB 361 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

QY 1047 TCCTT-----TAC 1054

DB 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

QY 1055 ATCAAAACCCACCTGTAGAGAGCTTTTGGCGTCTCTTAAATAGCTTGGCATGTCATT 1114

DB 401 IleLysProThrCysArgGluAlaPheAlaValSerLysIleAlaTrpHisValIle 420

QY 1115 AGGAATGTT 1123

DB 421 ArgAsnVal 423

RESULT 2

US-10-379-836-2

; Sequence 2, Application US/10379836

; Publication No. US20030215850A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON

; FILE REFERENCE: D0214NP

; CURRENT APPLICATION NUMBER: US/10/379, 836

; CURRENT FILING DATE: 2003-03-04

; PRIOR APPLICATION NUMBER: U.S. 60/361,523

; PRIOR FILING DATE: 2002-03-04

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 423

; TYPE: PRT

; ORGANISM: Papio hamadryas

US-10-379-836-2

Alignment Scores:

Pred. No.: 4,23e-170 Length: 423

Score: 1796.00 Matches: 347

Percent Similarity: 84.16% Conservative: 9

Best Local Similarity: 82.03% Mismatches: 13

Query Match: 64.67% Indels: 55

DB: 12 Gaps: 2

US-09-980-881A-1 (1-1573) x US-10-379-836-2 (1-423)

QY 18 ATGAAGCTTTGACGCTTGCAGTCTCTGTATCCCATTTCTCTCTCTGAGCAGCATGTC 77

DB 1 MetLysLeuCysSerLeuAlaValLeuValProLeuValLeuPheCysGluGlnHisVal 20

QY 78 TTCGCGTTCAGAGTGCACCAAGTTCTAGCTCTCTCTAGAACCTCTAGGCAAGTTCAA 137

DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40

QY 138 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGCGAGCCGGTAACAGCTGAC 197

DB 41 ValLeuGlnAsnLeuThrThrTyrGluValLeuTrpGlnProValThrAlaAsp 60

QY 198 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 257

DB 61 LeuLeuValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

QY 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAATGTCGACAAATGTG 317

DB 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

QY 318 CTTATTCAACGAGATTTCCACGACACACACTGAGCCCGAGCTCCGATCGTACTAT 377
 DB 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr 120
 QY 378 GAACAGTATCACTCACTAAATGAATCTATTCTGGATAGAAATTTATACTGAGAGCAT 437
 DB 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluLeuThrGluLysTyr 140
 QY 438 CTGTATATGCTTACAAAATCCCATTTGATTCCTCATTTGAGAAAGTACCCACTCTATGTT 497
 DB 141 ProAspMetLeuThrLysIleHisIleGlySerSerTyrGluLysHisProLeuTyrVal 160
 QY 498 TTAAGGTTTCTGAAAAGACAAACAGCCAAAATGCCATATGCTGATGCTGGAATC 557
 DB 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaMetTrpIleAspCysGlyIle 180
 QY 558 CATGCCAGAGAATGATCTCTCTGCTTCTGCTGTTGTTGTTGATAGGCCAT----- 608
 DB 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGlu 200
 QY 608 ----- 608
 DB 201 TyrTyrGlyIleIleGlyGluTyrThrAsnLeuLeuArgHisValAspPheTyrValMet 220
 QY 609 ----- 608
 DB 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
 QY 627 AACGGTTCTTCTATGGAACAATCATTTGCATCGAACAGACCTGATAGCACTTTGTC 686
 DB 241 AsnArgSerPheTyrAlaAsnAsnArgCysIleGlyThrAspLeuAsnArgAsnPheAla 260
 QY 687 TCAAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCATCTCGGAAACCTACTGTGGA 746
 DB 261 SerLysHisTrpCysGluGluGlyAlaSerPheSerCysSerGluThrTyrCysGly 280
 QY 747 CTTATCTCAGTCAGACAGAACAGAGTGAAGCAGTGTCTAGTTCCTGAGAAAGATATC 806
 DB 281 LeuTyrProGluSerGluProGluAlaLysAlaValAlaAsnPheLeuArgAsnIle 300
 QY 807 AACAGATTAAGCATACATCAGCATGCTCATCTACTCCAGCATATAGTGTTCATAT 866
 DB 301 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
 QY 867 TCCTATACAGAAAGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 926
 DB 321 SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 340
 QY 927 GTTCGTCTATTGACAAAATAGTAAATAACAGGTATACATAGGCCATGGCTCAGAA 986
 DB 341 ValArgAlaIleGlnLysThrSerLysAsnIleArgTyrThrHisGlyArgGlySerGlu 360
 QY 987 ACCTTATACCTAGCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAATAT 1046
 DB 361 ThrLeuTyrLeuAlaProGlyGlyAlaAspTrpIleTyrAspLeuGlyIleLysTyr 380
 QY 1047 TCGTT-----TAC 1054
 DB 381 SerPheThrIleGluLeuArgAspThrGlyLysTyrGlyPheLeuLeuProGluArgTyr 400
 QY 1055 ATCAAAACCCACTGTAGAGAAAGCTTTTGGCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
 DB 401 IleLysProThrCysLysAspAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420
 QY 1115 AGCAATGTT 1123
 DB 421 ArgAsnVal 423

RESULT 3

US-10-379-836-18

; Sequence 18, Application US/10379836

; Publication No. US20030215850A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
 ; FILE REFERENCE: D0214NP
 ; CURRENT APPLICATION NUMBER: US/10/379,836
 ; CURRENT FILING DATE: 2003-03-04
 ; PRIOR APPLICATION NUMBER: U.S. 60/361,523
 ; PRIOR FILING DATE: 2002-03-04
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 18
 ; LENGTH: 422
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-10-379-836-18

Alignment Scores:
 Pred. No.: 8,436-150 Length: 422
 Score: 1592.50 Matches: 309
 Percent Similarity: 79.15% Conservative: 25
 Best Local Similarity: 73.22% Mismatches: 33
 Query Match: 57.35% Indels: 56
 DB: 12 Gaps: 3

US-09-980-881A-1 (1-1573) x US-10-379-836-18 (1-422)

QY 18 ATGAAGCTTTGCAGCCTTCAGTCTTGTACCCATTTCTTCTCTGTGTGAGCAGCATGTC 77
 DB 1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleLeuTyr---GluGlnHisGly 19
 QY 78 TTCGCGTTTCAGAGTGGCAAGTTCCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 137
 DB 20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39
 QY 138 GTTCTACAGAATCTTACTACACATATGATGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
 DB 40 LeuLeuGlnAsnLeuThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59
 QY 198 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTCATGTGCGACAATGTG 257
 DB 60 PheIleGluLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79
 QY 258 AAGGCCATTTAAATGTGAGCGGAATTCATGCGAGTGTCTTGTGCGCAGACGCTGGAAGAT 317
 DB 80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99
 QY 318 CTTATTCAACAGCAGATTTCCACAGCAGACAGTCCAGCCCGCGCTCCGATCGTACTAT 377
 DB 100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119
 QY 378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 437
 DB 120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139
 QY 438 CTGTATATGCTTACAAAATCCCATTTGATTCCTCATTTGAGAAAGTACCCACTCTATGTT 497
 DB 140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159
 QY 498 TTAAGGTTTCTGAAAAGACAAACAGCCAAAATGCCATATGCTGATGCTGGAATC 557
 DB 160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179
 QY 558 CATGCCAGAGAATGATCTCTCTGCTTCTGCTGTTGTTGTTGATAGGCCAT----- 608
 DB 180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199
 QY 608 ----- 608
 DB 200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219
 QY 609 ----- 608
 DB 220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239


```
QY 363 TCGCATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTT 422
Db 115 Ser-----TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131
QY 423 ATAACTGAGAGGATCTCTGATGCTTTACAAAATCACAATTCGATGCTCTCTTGGAGAAG 482
Db 132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151
QY 483 TACCCACTCTATCTTTTAAAGTTTCTGGAAGAAGAACACACCCAAAATGCGATATGG 542
Db 152 AsnProLeuTyrValLeuLysIle-----GlyGluLysAsnGluArgArgLysAlaIlePhe 170
QY 543 ATTGACTGTGGAATCCATGCCAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 602
Db 171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190
QY 602 ----- 602
Db 191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210
QY 603 -----GGCCAT-----AAT 611
Db 211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230
QY 612 CGAATGTGGAAGAAACCGTTCTTCTATGCGAACAATCATTTGTCATCGGAACAGACCTG 671
Db 231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250
QY 672 AATGCAACTTTCTCTCAAACTGCTGTGAGGAAGGTGCATCTCAGTTCTCTCATCTCG 731
Db 251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269
QY 732 GAACACTGCTGCGATTTATCTGAGTCAGACCAAGAGTGAAGCAGTGGCTAGTTTC 791
Db 270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289
QY 792 TTGAGAAGAAATCAACAGATTAAGCATATCAATCAGCATGATTCATCATCTCCAGCAT 851
Db 290 IleArgSerHisLeuAsnGluLysValTyrIleThrPheHisSerTyrSerGlnMet 309
QY 852 ATAGTGTTCATTTCTTATACACGAAGTAAAGCAAGACCATGAGGAAGTCTCTCTA 911
Db 310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329
QY 912 GTAGCCAGTGAACAGTCTGCTGATATGACAAAATAGTAAATACCAAGTATACAT 971
Db 330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348
QY 972 GGCATGGCTCAGAAACCTTATACCTAGCTCTGAGAGGTGGGACGATGATCTATGAT 1031
Db 349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368
QY 1032 TTGGGCATCAATATTCTGTTTAC----- 1054
Db 369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388
QY 1055 -----ATCAACCCACCTGTAGAGAAGCTTTTCCCGCTGTCTCTAAATA 1099
Db 389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408
QY 1100 GCTTGGCATGTCATTAGGAAT 1120
Db 409 AlaLysTyrIleLeuLysHis 415
```

RESULT 7

```
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
```

```
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MPI01-156P1RM
; CURRENT APPLICATION NUMBER: US/10/229,546
```

```
; CURRENT FILING DATE: 2002-08-28
```

```
; PRIOR APPLICATION NUMBER: 60/316,575
```

```
; PRIOR FILING DATE: 2001-08-31
```

```
; NUMBER OF SEQ ID NOS: 9
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 2
```

```
; LENGTH: 417
```

```
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
```

```
; US-10-229-546-2
```

```
Alignment Scores:
```

```
Pred. No.: 1,55e-55 Length: 417
```

```
Score: 647.50 Matches: 143
```

```
Percent Similarity: 52.46% Conservative: 81
```

```
Best Local Similarity: 33.49% Mismatches: 132
```

```
Query Match: 23.32% Indels: 72
```

```
DB: 15 Gaps: 10
```

```
US-09-980-881A-1 (1-1573) x US-10-229-546-2 (1-417)
```

```
QY 6 ATGTCTGTGGATGAAGCTTTGCAGCTTGTGACCTCTGTACCCATTTGTTCTTCTGT 65
```

```
Db 5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal----- 19
```

```
QY 66 GAGCAGCATGTCTTCGCGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTCTAGAACCTCT 125
```

```
Db 20 -----ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34
```

```
QY 126 AGCAGATTCAGTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGCGACCG 185
```

```
Db 35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54
```

```
QY 186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTCTTAATGCATCTGAT 245
```

```
Db 55 GlyAlaThrHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74
```

```
QY 246 GTCGACAAATGTGAAGCCCATTTAAATGTGCGGAATTCATGCGAGTGTCTTGTGCGCA 305
```

```
Db 75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuHis 94
```

```
QY 306 GAGCTCGAAGATCTTATTCACAGCAGATTTCC---AACGACACATCAGCCCCAGCC 362
```

```
Db 95 AspLeuGlnGluLuleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114
```

```
QY 363 TCCGCATCGTACTGACAGATATCACTCACTAAATGAATCTATTCTTGGATAGAAATTT 422
```

```
Db 115 Ser-----TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131
```

```
QY 423 ATAACTGAGAGGATCTCTGATATGCTTACAAAATCCACATTCGATGCTCTCTTGGAGAAG 482
```

```
Db 132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151
```

```
QY 483 TACCCACTCTATCTTTTAAAGTTTCTGGAAGAAGAACACACCCAAAATGCGATATGG 542
```

```
Db 152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170
```

```
QY 543 ATTGACTGTGGAATCCATGCCAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 602
```

```
Db 171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190
```

```
QY 602 ----- 602
```

```
Db 191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210
```

```
QY 603 -----GGCCAT-----AAT 611
```

```
Db 211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230
```

```
QY 612 CGAATGTGGAAGAAACCGTTCTTCTATGCGAACAATCATTTGTCATCGGAACAGACCTG 671
```



```
Db 231 ArgMetTirpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250
QY 672 AATAGCAACTTTCTCTCCAAACACTGGTGTGAGGAAGGTGCATCTGCTCTCATCTCG 731
Db 251 AsnArgAsnPhe--AsnAlaSerTirpAsnSerIleProAsnThrAsnAspProCysAla 269
QY 732 GAACACTACTGTGACCTTTATCTGAGTCAGAACAGAGTGAAGCAGTGGCTAGTTTC 791
Db 270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289
QY 792 TTCAGAGAAATATCAACAGATTAAGCATATACATCAGCATGATCATCATCTCCAGCAT 851
Db 290 IleArgSerHisLeuAsnGluIleLysValTirleThrPheHisSerTyrSerGlnMet 309
QY 852 ATAGTGTTCCTATTCCTATATACACGAAGTAAAGCAAGCAACCATCAGGAACCTGTCTCTA 911
Db 310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329
QY 912 GTAGCAGTGAAGCAGTTCGTGCTATTGACAAAACCTAGTAAATAACAGGATATACACAT 971
Db 330 ValAlaLysIleGlyThrAspValLeu--SerThrArgTyrGluThrArgTyrIleTyr 348
QY 972 GGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGGTGGGACGATGGATCTATGAT 1031
Db 349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368
QY 1032 TTGGGCATCAAAATATTCGTTTAC-----1054
Db 369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388
QY 1055 -----ATCAAAACCCACCTGTAGAGAAAGCTTTTCCGCTGTCTCTAAATA 1099
Db 389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408
QY 1100 GCTTGGCATGTCTATTAGGAAT 1120
Db 409 AlaLysTyrIleLeuLysHis 415

RESULT 8
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; FILE REFERENCE: MP101-156PIRNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-546-9

Alignment Scores:
Pred. No.: 1,55e-55 Length: 417
Score: 647.50 Matches: 143
Percent Similarity: 52.46% Conservative: 81
Best Local Similarity: 33.49% Mismatches: 132
Query Match: 23.32% Indels: 72
DB: 15 Gaps: 10

US-09-980-881A-1 (1-1573) x US-10-229-546-9 (1-417)
QY 6 ATTGCTGTGGGATGAAGCTTTTGAGCCTTGCAGCTCTTGACCATTTGTTCTTCTGTG 65
```

```
Db 5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal-----19
QY 66 GAGCAGCATGTCTTCGCGTTTCCAGAGTGGCGGAATTCATCGAGTGTCTTCTAGCAACCTCT 125
Db 20 -----ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34
QY 126 AGCAAGTTCAGTCTTCTAAGAACTTCTACACATATGAGATTGTTCTCTGGCAGCCG 185
Db 35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTyrPro 54
QY 186 GTAACAGCTGACCTTATTCGTAAGAAAAACAAGTCCATTTTCTTAATGATCATCTGAT 245
Db 55 GlyAlaThrHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74
QY 246 GTCGCAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCATCGAGTGTCTTCTGGCGCA 305
Db 75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94
QY 306 GACGTGGAAGATCTTATTCACAGCAGAGATTTC---AACGACACATGACAGCCCCGAGCC 362
Db 95 AspLeuGlnGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114
QY 363 TCCGATCTGATCTACAGCAGTATCACTCACTAAATGAAATCTATCTTGGATAGATTT 422
Db 115 Ser-----TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131
QY 423 ATAACTGAGAGGATCTCTGATATGCTTACAAAAATCCACATTCGATCTCTATTTGAGAG 482
Db 132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151
QY 483 TACCCACTCTATGTTTAAAGGTTTCTGGAAAGAAACAAACAGCCAAATATGCATATGG 542
Db 152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgLysAlaIlePhe 170
QY 543 ATTGACTGTGAATCCATCCAGAGATGATCTCTCTGCTTCTCTGTTGTTGTTTCATA 602
Db 171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190
QY 602 -----602
Db 191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210
QY 603 -----GGCCAT-----AAT 611
Db 211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230
QY 612 CGAATGTGAGAAAGAACCGTTCTTCTATGCGAAACAATCATTTGCATTCGGAACAGACCTG 671
Db 231 ArgMetTirpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250
QY 672 AATAGCAACTTTCTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATCTCG 731
Db 251 AsnArgAsnPhe---AsnAlaSerTirpAsnSerIleProAsnThrAsnAspProCysAla 269
QY 732 GAAACCTACTGTGACCTTTATCTGAGTCAGAACAGAGTGAAGCAGTGGCTAGTTTC 791
Db 270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289
QY 792 TTGAGAGAAATATCAACAGATTAAGCATATACATCAGCATGATCATCTATCTATCTCCAGCAT 851
Db 290 IleArgSerHisLeuAsnGluIleLysValTirleThrPheHisSerTyrSerGlnMet 309
QY 852 ATAGTGTTCCTATTCCTATATACACGAAGTAAAGCAAGCAACCATCAGGAACCTGTCTCTA 911
Db 310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329
QY 912 GTAGCAGTGAAGCAGTTCGTGCTATTGACAAAACCTAGTAAATAACAGGATATACACAT 971
Db 330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348
QY 972 GGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGGTGGGACGATTTGATCTATGAT 1031
```

Db	349	GlyProIleGluSerThrIleTyrProIleSerGlySerLeuAspTrpAlaTyrAsp	368
Qy	1032	TTGGGCATCAAAATATCGTTTAC-----	1054
Db	369	LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu	388
Qy	1055	-----ATCAAAACCCACTGTAGAGAAGCTTTTGGCGTGTCTCTAAATA	1099
Db	389	LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle	408
Qy	1100	GCTTGGCATGTCATTAGGAAT	1120
Db	409	AlaLysTyrIleLeuLysHis	415

RESULT 9

```

US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-20

```

APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
APPLICANT: BURFORD, Neil; WALIA, Narinder K.
APPLICANT: LAL, Preeti G.; LEE, Sally
APPLICANT: TODD, Stephen; LO, Terence P.
APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
APPLICANT: AZIMZAI, Valda; LU, Yan
TITLE OF INVENTION: PROTEASES
FILE REFERENCE: PI-0167 USA
CURRENT APPLICATION NUMBER: US/10/274,639
CURRENT FILING DATE: 2002-10-18
PRIOR APPLICATION NUMBER: PCT/US01/22397
PRIOR FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: US 60/220,063
PRIOR FILING DATE: 2000-07-21
PRIOR APPLICATION NUMBER: US 60/221,680
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/223,544
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/224,717
PRIOR FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: US 60/225,988
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 60/227,568
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PERL Program
SEQ ID NO 12
LENGTH: 437
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

Alignment Scores:

Pred. No.: 2,48e-54 Length: 437
Score: 635.50 Matches: 147
Percent Similarity: 49.66% Conservative: 70
Best Local Similarity: 33.64% Mismatches: 147
Query Match: 22.88% Indels: 74
DB: 12 Gaps: 12

US-09-980-881A-1 (1-1573) x US-10-274-639-12 (1-437)

QY 9 GCTGTTGGGATGAAGCTTTCG-----AGCTTGGCAGTCTCTGTACCCATTGTTCTCTTC 62
DB 12 AlaAlaPheLeuProLeuGlySerValPheLeuLysIleLeuGlnPro----- 27
QY 63 TGTGAGCAGCATGCTTC-----GCGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCCT 116
DB 28 GlyHisSerHisLeuTyAsnAsnArgTyrAlaGlyAspLysValIleArgPheIlePro 47
QY 117 AGAACCTCTAGGCAAGTTCAGTCTTACAGAACTTCTACTACACATATGAGATGTTCTC 176
DB 48 LysThrGluGluGluAlaTyrAlaLeuLysIleSerTyrGlnLeuLysValAspLeu 67
QY 177 TGCAGCCGCTAACAGCTGACCTTATTGTCAAG-----AAAAACAAGTCCATTTTTT 230
DB 68 TrpGlnProSerSerIleSerTyrValSerGluGlyThrValThrAspValHisIlePro 87
QY 231 GTAATGTCATCTGATGTCGCAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCATGC 290
DB 88 GlnAsnGlySerArgAla-----LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyr 105
QY 291 AGTGTCTTGTGCGAGCTGGAAGATCTTATTCAACAGAGATTTCCACAGCACAGTC 350
DB 106 LysValLeuIleGluAspLeuLysThrLeuGluLysGlySerSerLeuHisThrGln 125
QY 351 AGCCCCGAGCTCCGATCG-----TACTATGACAGTATCACTCACTAAATGAATC 404
DB 126 ArgAsnArgArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIle 145

QY 405 TATTCTTGATAGAAATTTATTAAGGAGCATCTCTGATATGCTTACAAAAATCCACATT 464
DB 146 GlnAsnTrpMetHisHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIle 165
QY 465 GGATCTCTCAATTTGAGAAAGTACCACCTCTATGTTTAAAGGTTTCTGGAAGAACAAACA 524
DB 166 GlyArgSerTyrGluGlyArgSerLeuPheIleLeuLysLeu---GlyArgArgSerArg 184
QY 525 GCCAAAAATGCCATATGGAATGGAATCCATCCAGAGAAATGGATCTCTCTCTGCT 584
DB 185 LeuLysArgAlaValTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAla 204
QY 585 TTCTGCTTGTGTTTCATA----- 602
DB 205 PheCysGlnTrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArg 224
QY 602 ----- 602
DB 225 LysMetLeuAsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHis 244
QY 603 -----GGCCATTAATCGAATGTGGAAGAAAGACCGTTCTTTCTATGCGAAACATCAT 653
DB 245 PheSerTrpThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArg 264
QY 654 TGCATCGGAACAGACCTGATAGCAACTTGTCTCCAAACACTGTTGTGAGGAAGGTGCA 713
DB 265 CysArgGlyValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAla 283
QY 714 TCCAGTTCTCTCATGCTCGGAACCTACTGTGGAGTTTATCTCAGTCAGAACCCAGAGTG 773
DB 284 SerMetHisProCysAspSerThrTyrCysGlyProPheProGluSerGluProGluVal 303
QY 774 AAGCAGTGGCTAGTCTTCTTGAGAAAGAAATATCAACCAGATTAAGCATACATCAGCATG 833
DB 304 LysAlaValAlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPhe 323
QY 834 CATTCTATCTCCAGCATATAGTGTTCATATTTCTATACAGAAAGTAAAGCAAGAC 893
DB 324 HisAlaTyrAlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsn 343
QY 894 CATGAGGAAGTCTCTCTAGTACGCACTGAGCAGTTCGCTGCTATTGACAAAACCTAGTAA 953
DB 344 PheArgCysValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyr 362
QY 954 AATACAGAGTATACACATGCGCATGCGTCTCAGAAACCTTATACCTGCTCTCTGAGGTGG 1013
DB 363 GlyValArgTyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSer 382
QY 1014 GAGCATGGAATCTATGATTTGGGCATCAAAATATTCCTTTAC----- 1054
DB 383 MetAspTrpAlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThr 402
QY 1055 -----ATCAAAACCCACCTGTAGAGAGCTTTT 1081
DB 403 GlyTyrPheGlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMet 422
QY 1082 GCGCTGTCTCTAAATAGCTTGGCATGTCAATAGGAATGTTTAAATGCCCC 1132
DB 423 LeuAlaValLysAsnIleThrMetHisLeuLeuLysLys-----CysPro 437

RESULT 11

US-10-200-344-10
Sequence 10, Application US/10200344
Publication No. US20020173641A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Nehls, Michael
APPLICANT: Friedrich, Glenn
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
TITLE OF INVENTION: No. US20020173641A1e1 Human Carboxypeptidases and
; POLYNUCLEOTIDES Encoding the Same

```

; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 437
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-200-344-10

```

Best Local Similarity: 32.94% Mismatches: 143
Query Match: 22.85% Indels: 62
DB: 10 Gaps: 8

US-09-980-881A-1 (1-1573) x US-09-910-059-125 (1-716)

QY 24 CTTTGAGCCTTCAGTCTTGTGACCATTTGTTCTCTTCTGTGAGCAGCATGCTTCGCG 83
DB 1 MetLeuAlaLeuValLeuValThrValAlaLeuAlaSerAlaHisGlyGlu 20

QY 84 ---TTCAGAGTCCCAAGTTCTAGCTGCTCTTCTCTAGAACCTCTAGGCAAGTTCAAGTT 140
DB 21 HisPheGluGlyGluValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

QY 141 CTACAGAACTTCTACTACACATATGAGATTGTTCTCTGCGAGCCGGTAACAGCTGACCTT 200
DB 41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

QY 201 ATTGTGAAGAAAAACAGTCCATTTTGTGTAATGCATCTCATGTGCAATGTGAA 260
DB 61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

QY 261 GCCCATTTAAATGTGAGCGGAATTCATGCAGTGTCTTGTGCGAGAGTGAAGATCTT 320
DB 81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

QY 321 ATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCGCCGCTCCGCATCTACTATGAA 380
DB 101 ValGluAlaGlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGlu 117

QY 381 CAGTATCCTCACTAAATGAATCTATCTTGTGATAGAAATTTATATCTGAGAGCATCT 440
DB 118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnValAlaThrGluAsnPro 137

QY 441 GATATCTTCAAAATCCACATGTCATCTCTTGTGAGAGTACCCATCTATGTTTA 500
DB 138 AlaLeuIleSerArgSerValIleGlyThrPheGluGlyArgAlaIleTyrLeuLeu 157

QY 501 AAGTTTCTGAAAAGAACAAACAGACCAAAATGCCATATGCAATGCACTGTGGAATCCAT 560
DB 158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

QY 561 GCAGAGATGGATCTCTCTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 602
DB 177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

QY 602 -----GSCCAT----- 602

DB 197 TyrGlyArgGluIleGlnValThrGluLeuAspLysLeuAspPheTyrValLeuPro 216

QY 603 -----GSCCAT-----AATCGAATGTGAGAAAAGAAC 629

DB 217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpLysSerArgPheTrpArgLysThr 236

QY 630 CGTTCTTCTATCGCAACATCATCTGATCGGAACAGACCTGAATAGCACTTTGTCTCC 689

DB 237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

QY 690 AAACATGGTGTGAGAGGTGATCCAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 749

DB 256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275

QY 750 TATCTGTGTCAGAACAGAGTGAAGGCTAGTCTTCTGAGAGAAATATCAAC 809

DB 276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295

QY 810 CAGATTAAAGCATATCATGATGCTTCTACTCCAGCATATAGTGTGTTTCCATATTC 869

DB 296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315

QY 870 TATACAGAGTAAAGCAAGACCATGAGGAATCTGCTCTAGTACGCTGAGCAGTT 929

DB 316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335

QY 930 CGTGTCTATTGACAAACTAGTAAATAACAGGTATACATGCGCATGCTCGCAAAACC 989
DB 336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

QY 990 TTATACCTAGCTCCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATATTCG 1049
DB 355 IleTyrProSerAlaGlyThrSerLysAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374

QY 1050 TTATACATC----- 1057

DB 375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394

QY 1058 AAACCCACCTGTAGAGAACTTTTTCGCGCTGCTCTTAAATAGCTTGGCATGTCATTAGG 1117
DB 395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGlu 414

QY 1118 AATGTT 1123

DB 415 HisLeu 416

RESULT 13
US-09-910-059-113
; Sequence 113, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; TITLE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; PRIOR FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: humanised Fd mutant HCPB sequence
US-09-910-059-113

Alignment Scores:
Pred. No.: 6,64e-48 Length: 613
Score: 571.50 Matches: 136
Percent Similarity: 45.58% Conservative: 60
Best Local Similarity: 31.63% Mismatches: 118
Query Match: 20.58% Indels: 117
DB: 10 Gaps: 11

US-09-980-881A-1 (1-1573) x US-09-910-059-113 (1-613)

QY 20 GAAGCTTTTCAGCCTTGCAGTCTTGTACCCATTGTTCTTCTTCTGTGAGCAGCATGCTT 79
DB 238 GluLeuLysThrProLeuGlyAspThrThrHis-----ThrCysPro 251

QY 80 CGCGTTTCCAGAGTGGCAAGTTCTAGCTGCTCTTCTTAGAACCTCTAGGCAAGTTCAAGT 139
DB 252 ArgCysProGlu-----ProLysSerCys----- 259

QY 140 TCTACAGAACTTACTACAAATATAGATGTTTCTCTGCGAGCCCGGTAAACAGTCACT 199
DB 260 -----AspThrProPro 263

```
QY 200 TATTGTGAAGAAAAACAAGTCCATTTTGT-----AAATGC 238
Db 264 ProCysProArgCysProGluProLysSerCysAspThrProProCysProArgCys 283
QY 239 ATCTGATGTCGAATGTGAAGCCCAATTTAAATGTGAGCGGAATTCATGCAATGCTTT 298
Db 284 ProGluProLysSerCysAspThrProProCysProArgCysProAlaProGluLe 303
QY 299 GCTGGCAGAGTGAAGATCTTATTTCAACAGCAGATTTCACACGACACAGTACGCCCG 358
Db 303 uLeuGlyGly----- 306
QY 359 AGCTCCGCATCGTACTGAACAGATATCACTCAATAATGAATCATTTCTTTGGATAGA 418
Db 307 -AlaThrGlyHisSerTyrgluysrAsnLysrThrGluThrIleGluAlaThrGlu 326
QY 419 ATTATTAACCTGAGAGGCATCTGATATGCTTACAAAATCCACATTTGATCTCATTTGA 478
Db 326 nGlnValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGlyThrThrPheG 346
QY 479 GAAGTACCCACTCTATGTTTAAAGTTTCTGGAAGAACAAACACGCCAAAATGCCAT 538
Db 346 uGlyArgAlaIleTyrgluysrVal---GlyLysAlaGlyGlnAsnLysProAla 365
QY 539 ATGATTTGACTGTGGAATCCATCCAGAGATGATCTCTCTGCTTTCTGCTTTGGTT 598
Db 365 ePheMetAspCysGlyPheHisAlaArgGluThrIleSerProAlaPheCysGlnTrpPh 385
QY 599 CATA----- 602
Db 385 eValArgGluAlaValAlaArgThrTyrglyArgGluIleGlnValThrGluLeuAspLy 405
QY 603 -----GGCCAT----- 608
Db 405 sLeuAspPheTyrgluysrValLeuProValLeuAsnIleAspGlyTyrglyThrTrpThrLy 425
QY 609 -AATCGAATGTGAGAAAGACCGTTCTTCTATGCGAACCAATCATTCATCGCAAGACA 667
Db 425 sSerArgPheTrpArgLysThrArgSerThrHisThrGlySerSerCysIleGlyThrAs 445
QY 668 CCTGAATPAGCACTTTCTCTCAAAACACTGGTGTGAGGAAGGTGCATTCAGTTCTCCATG 727
Db 445 pProAsnArgAsnPhe---AspAlaGlyTrpCysGluIleGlyAlaSerArgAsnProCy 464
QY 728 CTGGAACCTACTGTGCACTTTATCTGTAGTCGAGACAGAGTGAAGCAGTGGCTAG 787
Db 464 sAspGluThrTyrglyProAlaAlaGluSerGluLysGluThrLysAlaLeuAlaAs 484
QY 788 TTTCTTGAGAGAAATATCAACAGATTAAAGCATACATCAGCATGCATTCATCTCCCA 847
Db 484 pPheIleArgAsnLysLeuSerSerIleLysAlaTyrgluysrThrIleHisSerTyrgly 504
QY 848 GCATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGACCATGAGGAACCTGTC 907
Db 504 nMetMetIleTyrglyProTyrglySerTyrglyLysLeuGlyGluAsnAlaGluLeuAs 524
QY 908 TCTAGTACCGAGTGAACAGTTCGTCTATGTATGACAAACTAGTAAATAATACAGGTATAC 967
Db 524 nAlaLeuAlaLysAlaThrValLysGluLeu---AlaSerLeuHisGlyThrLysTyrgly 543
QY 968 ACATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGCTGAGGTGGGACCATTTGGATCTA 1027
Db 543 rTyrglyProGlyAlaThrThrIleTyrglyProSerAlaGlyThrSerLysAspTrpAlaTy 563
QY 1028 TGATTTGGGCATCAATATTCGTTTACATC----- 1057
Db 563 rAspGlnGlyIleArgTyrglySerPheThrPheGluLeuArgAspThrGlyArgTyrglyPh 583
QY 1058 -----AAACCCACTGTAGAGAGCTTTTCCCGCTGCTCTAA 1095
Db 583 eLeuLeuProGluSerGlnIleArgAlaThrCysGluGluThrPheLeuAlaIleLysTy 603
```

```
QY 1096 AATAGCTTGGCATGTCTATTAGGAATGTT 1123
Db 603 rValAlaSerTyrgluysrValLeuGluHisLeu 612
```

RESULT 14

```
US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-61
```

Alignment Scores:

Pred. No.:	8,09e-48	Length:	374
Score:	570.00	Matches:	117
Percent Similarity:	52.47%	Conservative:	53
Best Local Similarity:	36.11%	Mismatches:	90
Query Match:	20.53%	Indels:	65
DB:	9	Gaps:	7

US-09-980-881A-1 (1-1573) x US-09-888-615-61 (1-374)

```
QY 324 CAACAGCAGATTCCAAACAGACAGTCAGGCCCGAGCCTCCGCATCGTAC---TATGAA 380
Db 29 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrglySerTyrgly 48
QY 381 CAGTATCACTCACTAAATGAATCTATTCTTGATAGAAATTTATACTAGAGGCATCCT 440
Db 49 IleTyrglyHisProMetGlyGluIleTyrglyTrpMetArgGluLysGlyLysTyrgly 68
QY 441 GATATGCTTACAAAAATCCACATTTGGATCCTCATTTGAGAAGTACCACCTCTATGTTTA 500
Db 69 GluValValThrGlnHisPheLeuGlyValThrTyrglyThrHisProMetTyrglyLeu 88
QY 501 AGGTTTCTGGAAGAACAAACAGCCAAATGCCATATGATGATGATGATGATGATGATGAT 560
Db 89 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 108
QY 561 GCCAGAGATGATCTCTCTCTGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
Db 109 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 128
QY 602 ----- 602
Db 129 HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrglyValLeuPro 148
QY 603 -----GGCCAT-----AATCGAATGTGAGAAAGAAC 629
Db 149 ValLeuAsnIleAspGlyTyrglyIleTyrglyThrThrAspArgLeuTrpArgLysSer 168
QY 630 CGTTCTTTCTATCGCAACATCATTCATCGGAGACAGCTGAATAGCACTTGTCTCC 689
Db 169 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 187
QY 690 AAACACTGTGTGAGGAAGGTGCATCCAGTTCCTCATGTCGGAACCTACTCTGCACTT 749
```



```

Db      378 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 397
Qy      930 CGTGCTATTGACAAACTAGTAAATAACACAGGTATACACATGCCCATGGCTCAGAAACC 989
Db      398 AsnAlaLeu--LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 416
Qy      990 TTATACCTAGCTCCTGGAGGTGGGACGATTGGATCTATGATTGGGCATCAAAATATTCG 1049
Db      417 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspIleGlyIleProPheSer 436
Qy      1050 TTTACA 1055
Db      437 TyrThr 438

```

Search completed: January 1, 2004, 19:25:30
 Job time : 98.6674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:25:26 ; Search time 30.8315 Seconds
(without alignments)
9812.904 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgtggatg.....aaaaaaaaaaaaaaaaaaaa 1573

Scoring table:

BXgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 segs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cpn2_1/uspto/spool/US09980881/runat_31122003_135749_24966/app_query.fasta_1.2958
-DB=pir_76 -QPMF=fastan -SUFFIX=tdr -MINMATCH=0 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CCN 1 1 71 @runat_31122003_135749_24966 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_76:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1885	67.9	423	2 A41204	carboxypeptidase B
2	695.5	25.0	415	2 A32129	carboxypeptidase B
3	655	23.6	417	1 A34487	carboxypeptidase A
4	647.5	23.3	417	2 A43929	carboxypeptidase A
5	634	22.8	416	1 A42332	carboxypeptidase B
6	589.5	21.2	306	1 CP80B	carboxypeptidase B
7	567.5	20.4	309	2 A38395	mast cell carboxyp
8	553	19.9	417	2 A32128	carboxypeptidase A
9	536	19.3	419	1 CP80A	carboxypeptidase A
10	529	19.0	419	1 CP8TA	carboxypeptidase A
11	508	18.3	417	2 A56171	carboxypeptidase A
12	489.5	17.6	419	1 S29127	carboxypeptidase A
13	437.5	15.8	528	2 T33527	hypothetical prote
14	400.5	14.4	667	2 T33526	hypothetical prote

15	397	14.3	323	2	T20507	hypothetical prote
16	391.5	14.1	527	2	T27572	hypothetical prote
17	369.5	13.3	303	1	CPCYB	carboxypeptidase B
18	362.5	13.1	455	2	T33572	hypothetical prote
19	358	12.9	666	2	T24170	hypothetical prote
20	299.5	10.8	430	2	S48976	ECM14 protein - ye
21	277.5	10.0	424	2	S17571	carboxypeptidase T
22	277	10.0	451	2	S20723	carboxypeptidase (
23	255.5	9.2	497	2	T40260	carboxypeptidase p
24	244.5	8.8	373	2	T26030	hypothetical prote
25	181.5	6.5	999	2	T36021	probable zinc-bind
26	135.5	4.9	1446	2	T30916	carboxypeptidase D
27	130.5	4.7	558	2	S75104	hypothetical prote
28	118.5	4.3	707	2	T40070	origin recognition
29	118	4.2	1216	2	H40629	hypothetical prote
30	112	4.0	1968	1	PN0093	genome polyprotein
31	105.5	3.8	985	2	D82776	pyruvate dehydroge
32	101.5	3.7	985	2	T29910	hypothetical prote
33	100.5	3.6	537	2	C64432	hypothetical prote
34	100	3.6	376	1	E69957	gamma-D-glutamyl-L
35	99.5	3.6	1389	2	I50090	carboxypeptidase g
36	99	3.6	1587	2	AB2012	hypothetical prote
37	98.5	3.5	606	2	E90536	hypothetical prote
38	98.5	3.5	983	2	B49284	immediate-early pr
39	98.5	3.5	1078	2	T44232	hypothetical prote
40	98.5	3.5	1160	2	I40589	parasporal crystal
41	98	3.5	491	2	B96739	hypothetical prote
42	97.5	3.5	355	2	C83850	gamma-D-glutamyl-L
43	97.5	3.5	7829	2	T15789	hypothetical prote
44	97	3.5	663	2	T37772	telomere length re
45	96	3.5	368	2	T21748	hypothetical prote

ALIGNMENTS

RESULT 1

A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N/Alternate names: plasma carboxypeptidase B
C/Species: Homo sapiens (man)
C/Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C/Accession: A41204
R/Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A/Title: Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase
A/Reference number: A41204; MUID:92042093; PMID:1939207
A/Accession: A41204
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-423 <EAT>
A/Cross-references: GB:M75106; NID:gl89686; PID:gl89687
C/Genetics:
A/Gene: GDB:CPB2
A/Cross-references: GDB:129546; OMIM:212070
A/Map position: 13q14.11-13q14.11
C/Superfamily: carboxypeptidase
C/Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F/1-22/Domain: signal sequence #status predicted <SIG>
F/23-114/Domain: activation peptide #status predicted <ACP>
F/115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Alignment Scores:
Pred. No.: 1,36e-158 Length: 423
Score: 1885.00 Matches: 365
Percent Similarity: 86.52% Conservatives: 1
Best Local Similarity: 86.29% Mismatches: 3
Query Match: 67.88% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-1 (1-1573) x A41204 (1-423)

QY 18 ATGAGAGCTTTCAGCGCTTCAGTTCCTTGTGACGAGCATGTC 77
|||||

Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 78 TTGCGGTTCAGAGTGGCCAAAGTTCTAGCTCTTCTTAGAACCTCTAGGCAAGTTCAA 137
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 138 GTTCTACAGAACTTACTACAACATATGAGATTGTTCTTGGCAGCGGTAACAGCTGAC 137
Db 41 ValLeuGlnAsnLeuThrThrThrTyrrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 198 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGAATGTG 257
Db 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 258 AAAGCCCATTAATGTGAGCGAATTCATGAGTCTTCTGTCGAGCGTGAAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY 318 CTTATTCAACAGCAGATTTTCCAAACGACACACAGTCAGCCCGAGCTCCGCATCGTACTAT 377
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrr 120
QY 378 GACAGTATCATCTCAATAAGAACTATCTTGGATAGAAATTAATACTGAGAGCAT 437
Db 121 GluGlnTyrrHisSerLeuAsnGluIleTyrrSerTrpIleGluPheIleThrGluArgHis 140
QY 438 CCTGATATGCTTACAAATCCCATGATGATCTCTCATTTGAGAGTACCCACTCTATGTT 497
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrrProLeuTyrrVal 160
QY 498 TTAAGGTTTCTGAAAGAACAACACAGCCCAAAATGCCATATGATGATCTGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 558 CATGCCAGAAATGGATCTCTCTGCTTCTGCTTCTGCTGTTTCATAGCCCAT----- 608
Db 181 HisAlaArgLutrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 608 ----- 608
Db 201 PheTyrrGlyIleIleGlyGlnTyrrThrAsnLeuLeuArgLeuValAspPheTyrrValMet 220
QY 609 -----AATCGAATGTGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyrrAspTyrrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 627 AACCGTTCTTTATGCGAACAAATCATGTCATCGAACACAGCTGAATAGCAACTTTGTC 696
Db 241 AsnArgSerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 687 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCATGTCGCGAAACCTACTGTGA 746
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrrCysGly 280
QY 747 CTTTATCTCAGTCAGACAGCAAGTGAAGCGAGTGGTAGTTTCTTGAGAACAAATATC 806
Db 281 LeuTyrrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 807 AACCAGATTAAAGCATACATCAGCATGCTATCTACTCCAGCATATAGTGTTCATAT 866
Db 301 AsnGlnIleLysAlaTyrrIleSerMetHisSerTyrrSerGlnHisIleValPheProTyrr 320
QY 867 TCCTATACAGAAAGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 926
Db 321 SerTyrrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
QY 927 GTTCGGCTATTACAAAACCTAGTAAATAACAGGTATACATCGCCATGGCTCAGAA 986
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrrThrHisGlyHisGlySerGlu 360
QY 987 ACCTTATACCTAGCTCTCTGGAGTGGGACGATTTGATCTATGATTTGGCGCATCAATAT 1046
Db 361 ThrLeuTyrrLeuAlaProGlyGlyAspAspTrpIleTyrrAspLeuGlyIleLysTyrr 380

QY 1047 TCCTT-----TAC 1054
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrrGlyPheLeuLeuProGluArgTyrr 400
QY 1055 ATCAAAACCCACTGTAGAGAAGCTTTTGGCGTGTCTCTAAATAGCTTGGCATGTCAAT 1114
Db 401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420
QY 1115 AGGAATGTT 1123
Db 421 ArgAsnVal 423
RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C:Accession: A32129; S17543
R:Clauuser, E.; Gardell, S.J.; Craik, C.S.; Macdonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A:Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparat
A:Reference number: A92693; MUID:89034324; PMID:3182872
A:Accession: A32129
A:Molecule type: DNA
A:Residues: 1-415 <CLA>
A:Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R:Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A:Title: Distribution of manganese in rat pancreas and identification of its primary bind
A:Reference number: S17543; MUID:91378950; PMID:1898371
A:Accession: S17543
A:Molecule type: protein
A:Residues: 109-116, 'X', 118-130 <KOD>
C:Genetics:
A:Gene: CPB
A:Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:354,376/Active site: Tyr, Glu #status predicted
Alignment Scores:
Pred. No.: 2,7e-53 Length: 415
Score: 695.50 Matches: 158
Percent Similarity: 52.74% Conservative: 63
Best Local Similarity: 37.71% Mismatches: 135
Query Match: 25.05% Indels: 64
DB: 2 Gaps: 8
US-09-980-881A-1 (1-1573) x A32129 (1-415)
QY 18 ATGAAGCTTTGACGCTTGCAGTCTTGCATTCCTCTCTCTGTGAGCAGCATGTC 77
Db 1 MetLeuLeuLeuLeuAlaLeuValSerValAlaLeuAlaHisAlaSerGluGluHis--- 19
QY 78 TTCGCGTTCCAGAGTGGCCAAAGTTCTAGCTGTCTCTCTAGAACCTCTAGGCAAGTTCAA 137
Db 20 -----PheAspGlyAsnArgValTyrrArgValSerValHisGlyGluAspHisValAsn 37
QY 138 GTTCTACACAACTTACTACACATATGAGATTGTTCTCTGGCAGCGCGTAAACAGCTGAC 197
Db 38 LeuIleGlnGluLeuAlaAsnThrLysGluIleAspPheTrpLysProAspAlaThr 57
QY 198 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACATGTG 257
Db 58 GlnValLysProLeuThrThrValAspPheHisValLysAlaGluAspValAlaAspVal 77
QY 258 AAAGCCCATTTAAATGTGAGCGAATTCATGAGTGTCTTCTGTCGACAGCTGGAAGAT 317
Db 78 GluAsnPheLeuGluGluAsnGluValHisTyrrGluValLeuIleSerAsnValArgAsn 97
QY 318 CTTATTCAACAGCAGATTTTCCAAACAGACACAGTCAGCCCGAGCTCCGCATCGTACTAT 377

Db 98 AlaLeuGluSerGlnPheAspSerHisThr-----ArgAlaSerGlyHisSerTyr 114
QY 378 GAACAGTATCCTACCTAAATGAATCTATTCTGTGATAGAAATTTATACTGAGAGCAT 437
Db 115 ThrLysTyrAsnLysTrpGluThrIleGluAlaTrpIleGlnValAlaThrAspAsn 134
QY 438 CTTGATATCTTACAAATCCACATGATCTCTCATTTGAGAGTACCCACTCTATGTT 497
Db 135 ProAspLeuValThrGlnSerValIleGlyThrThrPheGluGlyArgAsnMetTyrVal 154
QY 498 TTAAGGTTTCTGAAAGAACACAGCCAAATGCCATATGCTGATTCACCTGTGCAATC 557
Db 155 LeuLysIle---GlyLysThrArgProAsnLysProAlaIlePheIleAspCysGlyPhe 173
QY 558 CATGCCAGAGATGATCTCTCTGCTTCTGCTGCTGCTTCATA----- 602
Db 174 HisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArg 193
QY 602 ----- 602
Db 194 ThrTyrAsnGlnGluIleHisMetLysGlnLeuLeuAspGluLeuAspPheTyrValLeu 213
QY 603 -----GCCAT-----AATCGAATGTGAGAAAG 626
Db 214 ProValValAsnIleAspGlyTyrValTyrThrTrpThrLysAspArgMetTrpArgLys 233
QY 627 AACCGTTCTTCTATCGAACAATCATTCATCGACGACAGACCTGAATAGCACTTTGTC 686
Db 234 ThrArgSerThrMetAlaGlySerSerCysLeuGlyValArgProAsnArgAsnPhe--- 252
QY 687 TCMAACACTGTGTGAGGAGTGATCCAGTCTCTCATGCTCGGAAACCTACTGTGGA 746
Db 253 AsnAlaGlyTrpCysGluValGlyAlaSerArgSerProCysSerGluThrTyrCysGly 272
QY 747 CTTTATCTCTGAGTCAGAACAGAGTGAAGGAGTGTCTAGTTCTTGTGAGAGAAATATC 806
Db 273 ProAlaProGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLeu 292
QY 807 AACGAGTAAAGCATACATCAGCATGCTATCTACTCCAGCATATAGTGTTCATAT 866
Db 293 SerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetLeuTyrProTyr 312
QY 867 TCCTATACAGAAAGTAAAGCAAGACCATAGGAATCTGCTCTAGTACGAGTGAAGCA 926
Db 313 SerTyrAspTyrLysLeuProGluAsnTyrGluLeuAsnAlaLeuValLysGlyAla 332
QY 927 GTTCGCTATTGACAAACTAGTAAATACAGGTATACATACATGCGCATGCTCAGAA 986
Db 333 AlaLysGluLeu---AlaThrLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThr 351
QY 987 ACCTTATACCTAGCTCTGAGGTGGGAGCATTTGATCTATGATTTGGGCATCAATAT 1046
Db 352 ThrIleTyrProAlaAlaGlyLysSerAspAspTrpSerTyrAspGlnGlyIleLysTyr 371
QY 1047 TCCTTTACATC----- 1057
Db 372 SerPheThrPheGluLeuArgAspThrGlyPheGlyPheLeuLeuProGluSerGln 391
QY 1058 ---AAACCCAGCTAGAGAGCTTTTGGCGTCTCTAAATAGCTTGGCATGTC 1111
Db 392 IleArgGlnThrCysGluGluThrMetLeuAlaValLysTyrIleAlaAsnTyrVal 410

RESULT 3
A34487
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34487
R;Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A;Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member
A;Reference number: A34487; MUID:90062123; PMID:2584208
A;Accession: A34487

A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <REY>
C;Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:g309135
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F:356,378/Active site: Tyr, Glu #status predicted
Alignment Scores: 1.04e-49 Length: 417
Pred. No.: 655.00 Matches: 150
Score: 51.67% Conservative: 66
Percent Similarity: 35.89% Mismatches: 138
Best Local Similarity: 21.59% Indels: 65
Query Match: 1 Gaps: 9
DB: 1
US-09-980-881A-1 (1-1573) x A34487 (1-417)
QY 33 CTTGCGAGTCTGTGTACCCATTGTTCTTCTGTGAGCAGCATGTCTCGGTTCCAGAGT 92
Db 7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis-----PheAspArg 23
QY 93 GGCCAGTTCTAGCTGCTCTCTCTAGAACCTCTAGGCAAGTTCAAGTCTCTACAGATCTT 152
Db 24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43
QY 153 ACTACAACATATGAGATTGTTCTCTGGCAGCCGCTAACAGCTGACCTTATTGTGAAGAA 212
Db 44 ThrGlnSerIleGluLeuAspPheTyrTyrProAspAlaIleHisAspIleAlaValAsn 63
QY 213 AAACAAGTCCATTTTGTGTAATGATCTGATGTCGACATGTGAAGCCCTTTAAAT 272
Db 64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83
QY 273 GTGAGCGGAATCCATGCGAGTCTGCTGTCGCGACGCTGACAGCTTATTCAACAGCAG 332
Db 84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluLeuGluLysGln 103
QY 333 ATTTC---AAGCAGACAGTCAGCCCGCGAGCTCGCATCGTACTATGAACAGTAGTAC 389
Db 104 PheAspValLysAspGluIleAlaGlyArgHisSer-----TyrAlaLysTyrAsn 120
QY 390 TCCTAAATGAATCTATTCTTGGATAGAAATTTATTAAGTGAAGGATCTCTGATGCTT 449
Db 121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140
QY 450 ACAAAATCCATGATGCTCTCATTTGAGAGTACCCACTCTATCTTTTAAAGGTTTCT 509
Db 141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159
QY 510 GGAAGAAACAAACAGCCAAATGCCATATGATGACTGTGGAATCCATGCCAGAGAA 569
Db 160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179
QY 570 TGGATCTCTCTGCTTCTGCTTGTGCTTCATA-----GGCCAT 608
Db 180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199
QY 609 AAT----- 611
Db 200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219
QY 612 -----CGAATGTGGAAGAAACCCGTTCTTTC 638
Db 220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239
QY 639 TATGGAACAATCATTTGATCGGAACAGACCTGGAATAGCACTTTCTCTCCAAACACTGG 698
Db 240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258
QY 699 TGTGAGGAAGGTGCATCCAGTCTCTCATGCTCGGAAACCTACTGTGAGACTTTATCTGAG 758
Db 259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

QY 810 CAGATTAAAGCATACATCAGCATGCTTCTACTATCCAGCATATAGTGTTCATATCC 869
 Db 295 SerileylsalatyrLeuThrIleHisSerTyrSerGlnMetIleIleProTyrSer 314
 QY 870 TATACACGAAGTAAAGCAAGACCATGAGGAAGTGTCTCTAGTGGAGTGAAGCAGTT 929
 Db 315 TyrAlatyrLysLeuGlyGluAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 334
 QY 930 CGTGCTATTGACAAAAGTAAATACACAGGTATACATGCGCATGGCTCAGAAACC 989
 Db 335 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 353
 QY 990 TTATACCTAGCTCTGAGGTGGGAGGTATGGATCTATCATTTGGGCATCAATATTCG 1049
 Db 354 IleTyrProAlaAlaGlyGlySerAspAspIleAlaTyrAspGlnGlyIleArgTyrSer 373
 QY 1050 TTATCATC----- 1057
 Db 374 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 393
 QY 1058 AAACCCACCTGTAGAGAGCTTTGCGGTCTCTTAAATAGCTTGGCATGTCATTAGG 1117
 Db 394 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGlu 413
 QY 1118 AATGTT 1123
 Db 414 HisLeu 415
 RESULT 6
 CPOB
 carboxypeptidase B (EC 3.4.17.2) - bovine
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 02-Aug-1994
 C;Accession: A93797; A92150; A00912
 R;Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
 Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
 A;Title: Amino-acid sequence of bovine carboxypeptidase B.
 A;Reference number: A93797; PMID:75217824; PMID:1057162
 A;Accession: A93797
 A;Molecule type: protein
 A;Residues: 1-306 <TIT>
 R;Schmidt, J.J.; Hirs, C.H.W.
 J. Biol. Chem. 249, 3756-3764, 1974
 A;Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations of
 A;Reference number: A92150; PMID:74260705; PMID:4833744
 A;Accession: A92150
 A;Molecule type: protein
 A;Residues: 31-93;131-181;263-265;292-306 <SCH>
 R;Schmid, M.F.; Herriott, J.R.
 J. Mol. Biol. 103, 175-190, 1976
 A;Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
 A;Reference number: A92839; PMID:76265065; PMID:957425
 A;Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
 R;Plummer Jr., T.H.
 J. Biol. Chem. 244, 5246-5253, 1969
 A;Title: Isolation and sequence of peptides at the active center of bovine carboxypeptidase
 A;Reference number: A92051; PMID:70007159; PMID:5344132
 A;Contents: annotation; active site
 R;Kimmel, M.T.; Plummer Jr., T.H.
 J. Biol. Chem. 247, 7864-7869, 1972
 A;Title: Identification of a glutamic acid at the active center of bovine carboxypeptidase
 A;Reference number: A92119; PMID:73061487; PMID:4565668
 A;Contents: annotation; active site
 C;Superfamily: carboxypeptidase
 C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
 F;63-76,135-158,149-163/Disulfide bonds: #status experimental
 F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
 F;246,268/Active site: Tyr, Glu #status experimental
 Alignment Scores:
 Pred. No.: 6.3e-44 Length: 306
 Score: 589.50 Matches: 121

Percent Similarity: 56.58% Conservative: 51
 Best Local Similarity: 39.80% Mismatches: 75
 Query Match: 21.23% Indels: 58
 DB: 1 Gaps: 6
 US-09-980-881A-1 (1-1573) x CPOB (1-306)
 QY 375 TATGACAGTATCAGTACACATCAATGAATGAAATCTATTCTTGGATAGAAATTTATAACTGAGAGG 434
 Db 6 TyrGluLysTyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValAlaSerGlu 25
 QY 435 CATCTGATATGCTTCAAAAATCCACATTTGGATCTCTCATTTTGGAGAGTACCACCTCTAT 494
 Db 26 AsnProAspLeuIleSerArgSerAlaIleGlyThrPheLeuGlyAsnThrIleTyr 45
 QY 495 GTTTTAAAGTTCCTGGAAAAGAACAAACAGCCAAAATGCCATATGGATTGATCTGGA 554
 Db 46 LeuLeuLysVal---GlyLysProGlySerAsnLysProAlaValPheMetAspCysGly 64
 QY 555 ATCCATGCCAGAGAATGGATCTCTCTGCTTCTCTGCTTGTGGTTCATA----- 602
 Db 65 PheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaVal 84
 QY 602 ----- 602
 Db 85 ArgThrTyrArgGluIleHisMetThrGluPheLeuAspLysLeuAspPheTyrVal 104
 QY 603 -----GGCCAT-----AATCGAATGTGAGA 623
 Db 105 LeuProValValAsnIleAspGlyTyrIleTyrThrTrpThrThrAsnArgMetTyrArg 124
 QY 624 AAGAACCGTCTTCTATCGAACATCATTCGATCGGAACACACACCTGAATAGCAACTTT 683
 Db 125 LysThrArgSerThrArgAlaGlySerCysThrGlyThrAspLeuAsnArgAsnPhe 144
 QY 684 GTCTCCAAACACATGTGTGAGGAAGGTGATCCAGTCTCTCATGTCGGAACACTACTGT 743
 Db 145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163
 QY 744 GGACTTTATCTGAGTCAGAACACAGAAAGTGAAGGAGTGTGCTAGTTCTTTGAGAAGAAAT 803
 Db 164 GlySerAlaAlaGluSerGluLysGluSerLysAlaValAlaAspPheIleArgAsnHis 183
 QY 804 ATCAACACAGATTAAACATACATCAGCATGCTATCTATCCAGCATATAGTGTTCCTCA 863
 Db 184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrPro 203
 QY 864 TATCTCTATACACGAAGTAAAGCAAGACCATGAGGAAGTGTCTCTAGTAGCAGTGAA 923
 Db 204 TyrSerTyrAspTyrLysLeuProLysAsnAsnValGluLeuAsnThrLeuAlaLysGly 223
 QY 924 GCAGTTCGTCTATTGACAAAAGTAAATACACAGGTATACATGCGCATGGCTCA 983
 Db 224 AlaValLysLysLeu---AlaSerLeuHisGlyThrThrTyrSerTyrGlyProGlyAla 242
 QY 984 GAAACCTTATACCTAGCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAAA 1043
 Db 243 ThrThrIleTyrProAlaSerGlyGlySerAspAspIleAlaTyrAspGlnGlyLys 262
 QY 1044 TATTCGTTTACATC----- 1057
 Db 263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282
 QY 1058 -----AAACCCACCTGTAGAGAGCTTTGCGGTCTCTTAAATAGCTTGGCATGTC 1111
 Db 283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIleLysTyrValThrSerTyrVal 302
 QY 1112 AATAGGATGTT 1123
 Db 303 LeuGluHisLeu 306
 RESULT 7
 A38395

mast cell carboxypeptidase (EC 3.4.11.1) - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C/Accession: A38395; A31118
R/Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A/Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme activation by 1-phenyl-2-phenyl-1,3-bis(4-aminophenyl)propane
A/Reference number: A38395; MUID:91105153; PMID:1988052
A/Accession: A38395
A/Status: preliminary
A/Molecule type: protein
A/Residues: 1-309 <COL>
R/Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A/Reference number: A31118
A/Accession: A31118
A/Status: preliminary
A/Molecule type: protein
A/Residues: 2-309 <CO2>
C/Superfamily: carboxypeptidase
C/Keywords: hydrolase
P/248,270/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.: 5.59e-42 Length: 309
Score: 567.50 Matches: 120
Percent Similarity: 54.79% Conservative: 46
Best Local Similarity: 39.60% Mismatches: 80
Query Match: 20.44% Indels: 58
DB: 2 Gaps: 6

US-09-980-881A-1 (1-1573) x A38395 (1-309)

QY 375 TATGAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTTAACTGAGAGG 434
DB 375 TATGAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTTAACTGAGAGG 434
QY 435 CATCTCGATATGCTTACAAAATCCATTTGGATCTCTCATTTTGAGAAAGTACCCACTCTAT 494
DB 435 CATCTCGATATGCTTACAAAATCCATTTGGATCTCTCATTTTGAGAAAGTACCCACTCTAT 494
QY 495 GTTTTAAAGTTTCTGGAAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGA 554
DB 495 GTTTTAAAGTTTCTGGAAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGA 554
QY 48 ValLeuLysile---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66
DB 48 ValLeuLysile---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66
QY 555 ATCCATGCCAGAGAAATGATCTCTCTGCTTTCTGCTTGTGGTTCATA----- 602
DB 555 ATCCATGCCAGAGAAATGATCTCTCTGCTTTCTGCTTGTGGTTCATA----- 602
QY 67 IleHieAlaArgLutrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86
DB 67 IleHieAlaArgLutrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86
QY 603 -----GGCCATAAT----- 611
DB 603 -----GGCCATAAT----- 611
QY 87 LysSerTyrGlyLysAsnAlaIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106
DB 87 LysSerTyrGlyLysAsnAlaIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106
QY 612 -----CGAATGTCGAGA 623
DB 612 -----CGAATGTCGAGA 623
QY 107 LeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArg 126
DB 107 LeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArg 126
QY 624 AAGAACCGTTCTTTCTATGCGAACATCATTTGATCGGACACACCTGATAGCAACTTT 683
DB 624 AAGAACCGTTCTTTCTATGCGAACATCATTTGATCGGACACACCTGATAGCAACTTT 683
QY 127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146
DB 127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146
QY 684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATGCAGTTCCTCATGTCGCGAAACCTACTGT 743
DB 684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATGCAGTTCCTCATGTCGCGAAACCTACTGT 743
QY 147 ---AspValSerTrpAspSerProAsnThrAspAsnProCysLeuSerValTyrArg 165
DB 147 ---AspValSerTrpAspSerProAsnThrAspAsnProCysLeuSerValTyrArg 165
QY 744 GGACTTTATCTGAGTCAGAACACAGAGTGAAGCGAGTGGCTAGTTCTTTGAGAGAAAT 803
DB 744 GGACTTTATCTGAGTCAGAACACAGAGTGAAGCGAGTGGCTAGTTCTTTGAGAGAAAT 803
QY 166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185
DB 166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185
QY 804 ATCAACACAGATTAAACATACACAGATGATTCATCTCCAGCATATAGTGTTCCTCA 863
DB 804 ATCAACACAGATTAAACATACACAGATGATTCATCTCCAGCATATAGTGTTCCTCA 863
QY 186 LeuAsnSerIleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205
DB 186 LeuAsnSerIleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205

Qy	864	TATTCTCTATACGAAAGTAAAGCAAGACCATGAGGAACCTGCTCTCTAGTAGCCAGTGAA	923
Db	206	TyrGlyTyrThrIleLysLeuProProAsnHISGlnAspLeuLeuLysValAlaArgIle	225
Qy	924	GCAGTTCGTCTATTGACAAACATAGTAAAAAATACAGGTATACACATGCGCCATGGCTCA	983
Db	226	AlaThrAspValLeu---SerSerA:GTYrGLuThrArgTYrIleTYrGlyProlleAla	244
Qy	984	GAACACTTATACCTAGTCTCTCGAGGTGGGGACGATTGGATCTATGATTGGGCATCAAA	1043
Db	245	SerThrIleTYrLysThrSerGlySerSerLeuAspTrpAlaTYrAspLeuGlyIleLys	264
Qy	1044	TATTCGTTTAC-----	1054
Db	265	HISThrPheAlaPheGluLeuA:GspLysGlyLysSerGlyPheLeuLeuProGluSer	284
Qy	1055	--ATCAACCCACCTGTAGAGAGCTTTTGGCGCTGCTCTAAATAGCTTGGCATGTC	1111
Db	285	ArgIleLysProThrCysLysGluThrMetLeuSerValLysPheIleAlaLysTYrIle	3044
Qy	1112	ATTAGGAAT 1120	
Db	305	LeuLysHis 307	
RESULT 8			
A32128			
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat			
C:Species: Rattus norvegicus (Norway rat)			
C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999			
C/Accession: A32128			
R:Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf			
J. Biol. Chem. 263, 17828-17836, 1988			
A:Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning			
A:Reference number: A32128; MUID:89034323; PMID:3182871			
A:Accession: A32128			
A:Molecule type: mRNA			
A:Residues: 1-417 <AR>			
A:Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956			
C:Superfamily: carboxypeptidase			
C:Keywords: hydrolase; metallo-carboxypeptidase			
F:356,378/Active site: Tyr, Glu #status predicted			
Alignment Scores:			
Pred. No.: 1,11e-40 Length: 417			
Score: 553.00 Matches: 137			
Percent Similarity: 50.37% Conservative: 69			
Best Local Similarity: 33.50% Mismatches: 137			
Query Match: 19.91% Indels: 66			
DB: 2 Gaps: 13			
US-09-580-881A-1 (1-1573) x A32128 (1-417)			
Qy	18	ATGAAGCTTTGAGCCTTGAGTCCTTGATCCCATGTTCTCTCTGTGTAGCAGCATGTC	77
Db	1	MetArgLeuThrLeuLeuLeuAlaLeuLeuGlyTYrIleTYrCysGlnGlu-----	18
Qy	78	TTGCGCTTCAGAGTGGCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA	137
Db	19	---ThrPheValGlyAspGlnValLeuGluIleIleProSerHisGluGluGlnIleArg	37
Qy	138	GTTCCTACAGATCTTACTACAACA-----TATGAGATTGTTCTCTGGCAGCCGGTAACA	191
Db	38	ThrLeuLeuGlnLeuGluAlaGluGluHisLeuGluLeuAspPheTrpLysSerProThr	57
Qy	192	GCTGACCTTATTTGTGAAGAAAAAACAAGTCCATTTTTTGTAAATGCATGTGATGTCGAC	251
Db	58	-----IleProGlyGluThrValHisValArgValProPheAlaSerIleGln	73
Qy	252	AATGTGAAGCCCATTTAAATGTGACGGGAATTCACATGCTCTTGTGGCAGAGCTG	311
Db	74	AlaValLysValPheLeuGluSerGlnGlyIleAspTYrSerIleMetIleGluAspVal	93
Qy	312	GAAGATCTTATTCAACAGCAGCATTTTCCAAAC-----GACACAGTCAAGCCCCGAGCC	362

Db	94	GlnValLeuLeuAaspGlnGluArgGluMetLeuPheAenGlnGlnArgGluArgGly	113
Qy	363	TCCGCATCGTACTATGAACACTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTT	422
Db	114	GlyAenPheAenPheGluAlaTyRHisThrLeuGluGluIleTyRglnGluMetAaspAsn	133
Qy	423	ATAACTGAGAGGCATCTGTATGTTTACAAAAATCCACATTCGATCTCTCAATTTGAGAAG	482
Db	134	LeuValAlaGluAenProGlyLeuValSerLysValAenLeuGlySerSerPheGluAen	153
Qy	483	TACCCACTCTAATTTTAAAGGTTTCT--CGAAAGAACCAACAGCCAAAAATGCCATA	539
Db	154	ArgProMetAenValLeuLysPheSerThrGlyGlyAasp-----LysProAlaIle	170
Qy	540	TGGATTGACTGTGAATCCATCCAGACAGAAATGGATCTCTCTGCTTCTTCGTTGTGG---	596
Db	171	TrpLeuAaspAlaGlyIleHisAlaArgGluTrpValThrGlnAlaThrAlaLeuTrpThr	190
Qy	596	-----	596
Db	191	AlaAenLysIleAlaSerAaspTyR-GlyThrAaspProAlaIleThrSerLeuLeuAenThr	210
Qy	597	-----TTTCATA-----GGCCAT-----	608
Db	211	LeuAaspIlePheLeuLeuProValThrAasnProAaspGlyTyRValPheSerGlnThrThr	230
Qy	609	AATCGAATGGGAGAAAGAACCGTTCTTCTATCGGAACAATCATATGGATCGGACAGAC	668
Db	231	AsnArgMetTrpArgLysThrArgSerLysArgSerGlySerGlyCysValGlyValAasp	250
Qy	669	CTGAATAGCACTTTGTCTCCAAACACTGGTGTGAGGAAGTGATCCAGTTCCTCATGC	728
Db	251	ProAenArgAenTrp--AspAlaAenPheGlyGlyProGlyAlaSerSerSerProCys	269
Qy	729	TCGGAAACCTACTGTGGACTTTATCTTGAGTCAGAACCCAGAAAGTGAAGGAGTGGCTAGT	788
Db	270	SerAaspSerTyRHisGlyProLysProAasnSerGluValGluValLysSerIleValAasp	289
Qy	789	TTCTTGAGAAGAAATATCAACAGAGATTAAGACATACATCAGCATGCATTCATCTCCAG	848
Db	290	PheIleLysSerHis--GlyLysValLysAlaPheIleThrLeuHisSerTyR-SerGln	308
Qy	849	CATATAGTGTTCATATTCCTATTCACAGAAAGTAAAGCAAGACCATGAGGAACCTGTCT	908
Db	309	LeuLeuMetPheProTyRtyrGlyTyRlysCysThrLysProAaspAaspPheAenGluLeuAasp	328
Qy	909	CTAGTACGAGTGAGCAGTTCGTGCTATTACAAAACTAGTAAAAATACCAGGTATACA	968
Db	329	GluValAlaGlnLysAlaGlnAlaLeu---LysArgLeuHisGlySerTyRtyrLys	347
Qy	969	CATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACGATTGGATCTAT	1028
Db	348	ValGlyProIleCysSerValIleTyRglnAlaSerGlyGlySerIleAspTrpAlaTyR	367
Qy	1029	GATTGGGCATCAAATATTCGTTTTCATCAAC--	1061
Db	368	AspLeuGlyIleLysTyR-SerPheAlaPheGluLeuArgAaspThrAlaPheTyR-GlyPhe	387
Qy	1062	---CCACTGTAGAGAAGCTTTTGGCG	1085
Db	388	LeuLeuProAlaLysGlnIleLeuPro	396

RESULT 9

КРБОВ
СРБОВА

carboxypeptidase A (EC 3.4.17.1) precursor [validated] - bovine

N;Alternate names: procarboxypeptidase A activation peptide

C/Species: *Bos primigenius taurus* (cattle)

C;Date: 24-Apr-1984 #sequence revision 30-Jun-1992 #text change 05-May-2000

C;Accession: JN0126; JT0440; A31406; S00059; A388

R;Le Hueerou, I.; Guilloteau, P.; Toullec, R.; Pu

A:Title: Cloning and nucleotide sequence of a bovine pancreatic procarboxypeptidase A

A;Reference number: JN0126; MUID:91151335; PMID:1998496
A;Accession: JN0126
A;Molecule type: mRNA
A;Residues: 1-419 <LEH>
A;Cross-references: GB:M61851; NID:g162788; PIDN:AAA30426.1; PID:g162789
R;Fogliizzo, E.; Bonicel, J.; Kerfelec, B.; Granon, S.; Chapus, C.
Biochim. Biophys. Acta 954, 183-188, 1988
A;Title: Primary structure of the activation peptide from bovine pancreatic procarboxypeptidase A
A;Reference number: JT0440; MUID:88209583; PMID:3365436
A;Accession: JT0440
A;Molecule type: protein
A;Residues: 17-110 <FOG>
R;Wade, R.D.; Hass, G.M.; Kumar, S.; Walsh, K.A.; Neurath, H.
Biochimie 70, 1137-1142, 1988
A;Title: The amino acid sequence of the activation peptide of bovine pro-carboxypeptidase A
A;Reference number: A31406; MUID:89150306; PMID:3147705
A;Accession: A31406
A;Molecule type: protein
A;Residues: 17-110 <WAD>
R;Chapus, C.; Kerfelec, B.; Foglizzo, E.; Bonicel, J.
Eur. J. Biochem. 166, 379-385, 1987
A;Title: Further studies on the activation of bovine pancreatic procarboxypeptidase A by
A;Reference number: S00059; MUID:87275921; PMID:3609014
A;Accession: S00059
A;Molecule type: protein
A;Residues: 17-69 <CHA>
R;Bradshaw, R.A.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 63, 1389-1394, 1969
A;Title: The amino acid sequence of bovine carboxypeptidase A.
A;Reference number: A38834; MUID:70050818; PMID:5260942
A;Accession: A38834
A;Molecule type: protein
A;Residues: 111-137,'Q',139-140,'E',142-417 <BRA>
R;Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 961-972, 1971
A;Title: Amino acid sequence of bovine carboxypeptidase A. Isolation and characterization
A;Reference number: A90360; MUID:71155185; PMID:4927805
A;Accession: A90360
A;Molecule type: protein
A;Residues: 214-411 <BR2>
R;Petra, P.H.; Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 8, 2762-2768, 1969
A;Title: Identification of the amino acid replacements characterizing the allotropic form
A;Reference number: A90556; MUID:63283620; PMID:5817619
A;Contents: annotation; allelic variant
A;Note: an allelic variant having 179-Val, 228-Glu, and 305-Leu occurs with nearly equal
R;Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A;Reference number: A50963; PDB:4CPA
A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137,'Q',139-141
R;Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 168, 367-387, 1983
A;Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A;Reference number: A92902; MUID:83294519; PMID:6887246
A;Contents: annotation; X-ray crystallography, 1.54 angstroms
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <Sig>
F;17-110/Domain: activation peptide #status experimental <ACP>
F;111-419/Product: carboxypeptidase A #status experimental <MAT>
E;179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F;248-271/bisulfide bonds: #status experimental
F;358,380/Active site: Tyr, Glu #status experimental

Alignment Scores:

Qy	843	TCCAGCATATAGTGTTCATATCTCATTATACAGAAAGTAAGAAGCAACATGAGGAA	902
Db	309	SerGlnLeuLeuLeuTyProTyrGlyTyrThrGlnSerIleProAspLysThrGlu	328
Qy	903	CTGCTCTAGTAGCCAGTGAAGCAGTTTCGGTGCTATTGACAAAACACTAGTAAAAAATACCAGG	962
Db	329	LeuAsnGlnValAlaLysSerAlaValGluAlaLeu---LysSerLeuTyrGlyThrSer	347
Qy	963	TATACATCGGCATGGCTCAGAAAACCTTATACCTAGCTCCTCGAGGTGGGACGATTGG	1022
Db	348	TyrLysTyrGlySerIlelleThrThrIleTyrGlnAlaSerGlyGlySerIleAspTrp	367
Qy	1023	ATCATGANTTTGGGCATCAAATAATTCGTTTACATCAAA-----	1060
Db	368	SerTyrAsnGlnGlyIleLysTyrSerPheThrPheGluLeuArgAspThrGlyArgTyr	387
Qy	1061	-----CCCACCTGTAGACAAGCTTTTGGCGCTGTC	1090
Db	388	GlyPheLeuLeuProAlaSerGlnIlelleProThrAlaGlnGluThrTrpLeuGlyVal	407
Qy	1091	TCTAAATAGCTTGGCATGTCATTAGGAATGTT	1123
Db	408	LeuThrIleMetGluHisThrLeuAsnAsnLeu	418

RESULT 10
CPRTA

```

carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standing, D.N.; Pi
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characteri
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A:Molecule type: mRNA
A;Residues: 1-419 <OU>
A;Cross-references: GBJ:J00713; NID:G203364; PIDN:AAA40893.1; PID:G203365
R;Clauser, E.; Gardelli, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.O.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A:Molecule type: DNA
A;Residues: 1-260, 'F', 262, 'M', 264-346, 'K', 348-419 <CLA>
C;Genetics:
A;Gene: CPRA1
A;Intons: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A;Note: the authors translated the codon GTC for residue 196 as Ile
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digest
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <PRO>
F;111-419/Product: carboxypeptidase A #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

```

Alignment Scores:			
Pred. No.:	1,48e-38	Length:	419
Score:	529,00	Matches:	141
Percent Similarity:	47,94%	Conservative:	68
Best Local Similarity:	32,34%	Mismatches:	151
Query Match:	19,105%	Indels:	76
DB:	1	Gaps:	13

US-09-980-881A-1 (1-1573) vs CPRTA (1-419)

Qy	21	AAGCTTTGACGCTTCAGTCCCTTGATCCCATGTTCTCTCTCTGTGAGCAGCATGCTTC	80
Db	3	ArgLeuLeulleuSerLeuLeuGluAlaValCysGlyAsnGluAsn-----	19

Qy	843	TCCAGCATATAGTGTTCATATCTCATTATACAGAAAGTAAGAAGCAACATGAGGAA	902
Db	309	SerGlnLeuLeuLeuTyrrProTyrGlyTyrThrGlnSerIleProAspLysThrGlu	328
Qy	903	CTGCTCTAGTAGCCAGTGAAGCAGTTTCGGTGCTATTGACAAAACACTAGTAAAAATACCAGG	962
Db	329	LeuAsnGlnValAlaLysSerAlaValGluAlaLeu---LysSerLeuTyrGlyThrSer	347
Qy	963	TATACATCGGCATGGCTCAGAAAACCTTATACCTAGCTCCTCGAGGTGGGACGATTGG	1022
Db	348	TyrLysTyrGlySerIlelleThrThrIleTyrGlnAlaSerGlyGlySerIleAspTrp	367
Qy	1023	ATCATGANTTTGGGCATCAAATAATTCGTTTACATCAAA-----	1060
Db	368	SerTyrAsnGlnGlyIleLysTyrSerPheThrPheGluLeuArgAspThrGlyArgTyr	387
Qy	1061	-----CCCACCTGTAGACAAGCTTTTGGCGCTGTC	1090
Db	388	GlyPheLeuLeuProAlaSerGlnIlelleProThrAlaGlnGluThrTrpLeuGlyVal	407
Qy	1091	TCTAAATAGCTTGGCATGTCATTAGGAATGTT	1123
Db	408	LeuThrIleMetGluHisThrLeuAsnAsnLeu	418

RESULT 10
CPRTA

```

carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standing, D.N.; Pi
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characteri
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A;Molecule type: mRNA
A;Residues: 1-419 <OU>
A;Cross-references: GB|J00713; NID:G203364; PIDN:AAA40893.1; PID:G203365
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.O.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A;Molecule type: DNA
A;Residues: 1-260, 'F', 262, 'M', 264-346, 'K', 348-419 <CLA>
C;Genetics:
A;Gene: CPA1
A;Intons: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A;Note: the authors translated the codon GTC for residue 196 as Ile
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digest
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <PRO>
F;111-419/Product: carboxypeptidase A #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

```

Alignment Scores:			
Pred. No.:	1,48e-38	Length:	419
Score:	529,00	Matches:	141
Percent Similarity:	47,94%	Conservative:	68
Best Local Similarity:	32,34%	Mismatches:	151
Query Match:	19,105%	Indels:	76
DB:	1	Gaps:	13

US-09-980-881A-1 (1-1573) vs CPRTA (1-419)

Qy	21	AAGCTTTGACGCTTCAGTCCCTTGATCCCATGTTCTCTCTCTGTGAGCAGCATGCTTC	80
Db	3	ArgLeuLeulleuSerLeuLeuGluAlaValCysGlyAsnGluAsn-----	19

```
Db      20  ---PheValGlyHisGlnValLeuArgIleSerAlaAlaAspGluAlaGlnValGlnLys 38
QY      141 CTACAGAAATCTTACTACACATATGAGATTGTTCTC-----TGGCAGCCGGTAACAGCT 194
Db      39 ValysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgAspAlaAlaArg 58
QY      195 GACCTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTGCGCAAT 254
Db      59 AlaGlyIleProIleAspValArgValProPhe-----ProSerIleGlnSer 74
QY      255 GTGAAGCCCATTTAAATGTGAGCGGAATTCAGTGTCTTGTGGCAGCGTGGAA 314
Db      75 ValysAlaPheLeuGluTrpHisGlyIleSerTrpGluIleMetIleGluAspValGln 94
QY      315 GATCTTATT-----CAACAGCAGATTTCCAAACACACAGCAGTCCAGCCCGAGCTCC 365
Db      95 LeuLeuLeuAspGluGluLysGlnHisMetSerAlaPheGlnAlaArgAlaLeuSerThr 114
QY      366 GCATCGTAC---TATGAACAGTATCACTCAATAATGAAATCTATTCTTGGATAGAATTT 422
Db      115 AspSerPheAsnTrpAlaThrTrpHisThrLeuAspGluIleTrpGluPheMetAspLeu 134
QY      423 ATAACTGAGAGGCATCTGTATATGCTTTACAAAATCCACATTCGATCCTCATTTGAGAAG 482
Db      135 LeuValAlaGluHisProGlnLeuValSerLysIleGlnIleGlyAsnThrPheGluGly 154
QY      483 TACCACTCTATGTTTAAAGTTTCTGGAAGAAACAACAGCCAAAATGCGATATGG 542
Db      155 ArgProIleHisValLeuLysPheSer---ThrGlyGlyThrAsnArgProAlaIleTrp 173
QY      543 ATTGACTGTGGAATCCATGCCAGAGAATGATCTCTCTGCTTCTGCTTGTGTTTCATA 602
Db      174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193
QY      603 GGCCAT----- 608
Db      194 LysLysValThrLysAspTrpGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213
QY      609 -----AAT 611
Db      214 AspilePheLeuGluIleValThrAsnProAspGlyPheAlaTrpThrHisLysThrAsn 233
QY      612 CGAATGTGGAAGAAAGACCTGTTCTTCTATGCGAACAAATCAATTCGATCGGAACAGACTG 671
Db      234 ArgMetTrpArgLysThrArgSerHisThrGlnGlySerLeuCysValGlyValAspPro 253
QY      672 AATAGCAACTTTGTCTCCAAACACTGG-----TGTGAGGAAGTGTGATCCAGT 719
Db      254 AsnArgAsn-----TrpAspAlaGlyLeuGlyLysAlaGlyAlaSerSer 268
QY      720 TCCTCATGTCGGAACCTACTGTGGACTTTATCTGAGTCAGACAGACAGAGTGAAGCA 779
Db      269 AsnProCysSerGluThrTrpArgGlyLysPheProAsnSerGluValGluValLysSer 288
QY      780 GTGCTAGTTCTTGTGAAGAAATATCAACAGATATAAGCATATCAGCATGCAATTC 839
Db      289 IleValAspPheValThrSerHisGlyAsn---IleLysAlaPheIleSerIleHisSer 307
QY      840 TACTCCAGCATATAGTGTGTTTCCATATTCCTATACAGAAAGTAAAGCAAGCATTGAG 899
Db      308 TyrSerGlnLeuLeuLeuTrpProTyrGlyTrpThrSerGluProAlaProAspGlnAla 327
QY      900 GAAGTGTCTAGTAGCAGTGAAGCAGTTCGTGCTATTGACAAAACATGATAAAATACC 959
Db      328 GluLeuAspGlnLeuAlaLysSerAlaValThrAlaLeu---ThrSerLeuHisGlyThr 346
QY      960 AGGTATACATGCGCCATGGCTCAGAAACCTTATACCTAGCTCTGAGGTGGGAGCAT 1019
Db      347 GluPheLysTrpGlySerIleIleAspThrIleTrpGlnAlaSerGlySerThrIleAsp 366
QY      1020 TGGATCTATGATTGGGCATCAATATATCGTTTATCATCAAC----- 1061
```

```
Db      367 TrpThrTrpSerGlnGlyIleLysTrpSerPheThrPheGluLeuArgAspThrGlyLeu 386
QY      1062 -----CCACTGTAGAGAAGCTTTTCCGCTGTCTCTAAATAGCTTGGCATG 1109
Db      387 ArgGlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGluGlu-----Thr 403
QY      1110 TCATTAGGAATGTTTAAATGCCCTGATTTTATCATTTCTGCTTCCGTAT 1157
Db      404 TrpLeuAlaLeuLeuThrIleMetAspHisThrValLysHisProTyr 419

RESULT 11
A56171
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence revision 11-Aug-1995 #text_change 22-Jun-1999
C;Accession: A56171; S02809; S71395
R;Catalase, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:g790226; PIDN:AAV4425.1; PID:g790227
A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: S',18-36,N',38-43 <PAS>
R;Jaethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton,
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 ar
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-301,'T',303-338,'SR',341-417 <LA>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;11-417/Product: activation peptide #status predicted <ACP>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.: 1,07e-36 Length: 417
Score: 508.00 Matches: 128
Percent Similarity: 48.48% Conservative: 63
Best Local Similarity: 32.49% Mismatches: 137
Query Match: 18.29% Indels: 66
DB: 2 Gaps: 11

US-09-980-881A-1 (1-1573) x A56171 (1-417)
QY 72 CATGCTCTCTC-----CGGTTCCAGAGTGGCCAAAGTTCTAGCTGTCTCTCTAGAAC 122
Db 13 HisIleTrpCysLeuGluThrPheValGlyAspGlnValLeuGluIleValProSerAsn 32
QY 123 TCTAGGCAAGTTCAGTTCTACAGATCTTACTACACATATGAGATTGTTCTC----- 176
Db 33 GluGluGlnIleLysLeuLeuGlnLeuGluAlaGlnGlnHisLeuGlnLeuAspPhe 52
```

Qy	177	TGGCAGCCGGTAAACAGCTGACCTTATTGTGTGAAGAAAAAACAAAGTCCATTTTTTTTGTAAAT	236
Db	53	TrpLysSerProThr-----ThrProGlyGluThrAlaHisValArgValPro	68
Qy	237	GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGACGGCAATTCATGACGATGTC	296
Db	69	PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaTySerIle	88
Qy	297	TTGTGTCGACGAGTGGGAAGATCTTTATTCAACACAGCAGAGATTTTCCAAACGACACACAGTGCAGCCC	356
Db	89	MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg	108
Qy	357	CGAGCCTCCGATCG-----TACTATGAACAGTATCACTCACTAAATGAATCTAT	407
Db	109	ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaTyHisThrLeuGluGluIleSer	128
Qy	408	TCTTGGATAGAATTTATAACTGACGAGCATCCTGATATGCTTACAAAATCCATTTGGA	467
Db	129	GlnGluMetAspAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly	148
Qy	468	TCCTCATTTGAGAAAGTACCCACTCTATGTTTTTAAAGGTTTCT---GGAAAAAGAACAAACA	524
Db	149	SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyGlyAsp	166
Qy	525	GCCAAAAATGCCATATGATGACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCT	584
Db	167	--LysProAlaIleIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla	185
Qy	585	TTCTGCTTGTGGTTCATAGGC-----	605
Db	186	ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr	205
Qy	605	-----	605
Db	206	SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal	225
Qy	606	-----CATATCGAATGTGGAGAAAGACCGTCTTCTTCTATGCGAACAAATCAT	653
Db	226	PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLysValSerGlySerLeu	245
Qy	654	TGCATCGGAACAGACTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA	713
Db	246	CysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla	264
Qy	714	TCCAGTTTCTCATCTCGGAAACCTACTGTGGACTTATCTCTGAGTCCAGACCAAGATG	773
Db	265	SerSerAsnProCysSerAspSerTyrHisGlyProSerAlaAsnSerGluValGluVal	284
Qy	774	AAGCAGCTGGCTAGTCTTCTTGAGAAGAAATATCAACACAGATTAAGCATATCATCAGATG	833
Db	285	LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleIleLeu	303
Qy	834	CATTCTATCTCCAGCATATAGTGTTCCTATTCCTATATACAGAAATPAAAGCAAGAC	893
Db	304	HisSerTyrSerGlnLeuLeuMetPheProTyrGlyTyrLysCysThrLysLeuAspAsp	323
Qy	894	CATGAGAACTGTCTTAGTAGCCAGCTGAACGAGTTCGTGTATGTACAAAATAGTAAA	953
Db	324	PheAspGluLeuSerGluValAlaIleGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis	342
Qy	954	AATACCAAGGTATACATACATGGCCATGGCTCGAAGACCTTATACCTAGTCTCTGGAGGTGGG	1013
Db	343	GlyThrLysTyrLysValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySer	362
Qy	1014	GACGATGGATCTATGATTTGGGCATCAAAATATTCTGTTTACATCAAAAC-----	1061
Db	363	IleAspTrpSerTyrAspTyrGlyIleLysTyrSerPheAlaPheGluLeuArgAspThr	382
Qy	1062	-----CCACCTGTAGAGAAGCTTTTGGCG	1085
Db	383	GlyArgTyrGlyPheLeuLeuProAlaArgGlnIleLeuPro	396

RESULT 12

S29127
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N:Alternate names: pancreatic carboxypeptidase A1
C:Species: Homo sapiens (man)
C:Date: 25-Feb-1994 #sequence revision 19-Jan-1996 #text change 18-Jun-1999
C:Accession: S29127; A34205; S08253; S02810; S71394; S02811
R:Catapus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver, Biochem. J. 287, 299-303, 1992
A:Title: cDNA cloning and sequence analysis of human pancreatic procatboxypeptidase A1.
A:Reference number: S29127; MUID:93038569; PMID:1417781
A:Accession: S29127
A:Molecule type: mRNA
A:Residues: 1-419 <CAT>
A:Cross-references: EMBL:X67318; NID:G35329; PIDN:CAA47732.1; PID:G35330
R:Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A:Title: Human carboxypeptidase A identifies a BgIII RFLP and maps to 7q31-qter.
A:Reference number: A34205; MUID:90196012; PMID:1969228
A:Accession: A34205
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 330-396 <STE>
A:Note: the authors translated the codon CTG for residue 391 as Val
R:Moulard, M.; Michon, I.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A:Title: Further studies on the human pancreatic binary complexes involving procarboxypeptidase A1.
A:Reference number: S08253; MUID:90169111; PMID:2307232
A:Accession: S08253
A:Molecule type: protein
A:Residues: 17-43; XXX', 114-135 <MOU>
R:Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A:Title: Purification and properties of five different forms of human procarboxypeptidase A1.
A:Reference number: S02809; MUID:89153096; PMID:2920728
A:Accession: S02810
A:Molecule type: protein
A:Residues: 17-42 <PAS>
R:Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton, Arch. Biochem. Biophys. 332, 8-18, 1996
A:Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 ar
A:Reference number: S71394; MUID:96400327; PMID:8806703
A:Accession: S71394
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-419 <LAE>
C:Genetics:
A:Gene: GDB:CPA1; CPA
A:Cross-references: GDB:120597; OMIM:114850
A:Map position: 7q32-7qter
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status predicted <ACP>
F:111-419/product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F:179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F:248-271/Disulfide bonds: #status predicted
F:358,380/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.: 4,66e-35 Length: 419
Score: 489.50 Matches: 131
Percent Similarity: 47.63% Conservative: 60
Best Local Similarity: 32.67% Mismatches: 127
Query Match: 17.63% Indels: 83
DB: 1 Gaps: 13

US-09-980-881A-1 (1-1573) x S29127 (1-419)

Qy 24 CTTTGCAGCTTGCGAGTTCCTTGACCCATTGTTCTCTTGAGCAGCATGCTTCGCG 83
||| |||||:|||||:|||||
Db 4 LeuLeuValLeuSerValLeuLeuGlyAlaValPheGlyLysGluasp----- 19
Qy 84 TTCCAGAGTGGCCAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAGATTCTA 143

```
Db 20 PheValGlyHisGlnValLeuArgIleSerValAlaAspGluAlaGlnValGlnLeuVal 39
QY 144 CAGAACTCTTACTCAACATATGAGATTGTTCTC-----TGGCAG----- 182
Db 40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProAlaHisPro 59
QY 183 -----CCGGTAACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTTTGTAAAT 236
Db 60 GlySerProIleAspVal-----ArgValProPhe----- 69
QY 237 GCATCTGTCGACCAATGTGAAGCCCATTTAAATGTGAGCGCAATTCATCGAGTGC 296
Db 70 ---ProSerIleGlnAlaValIlePheLeuGluSerHisGlyIleSerTyrGluThr 88
QY 297 TTGCTGCGACAGCTGGAGATCTTATT-----CAACAGCAGATT----- 335
Db 89 MetIleGluAspValGlnSerLeuLeuAspGluGlnGlnMetPheAlaPheArg 108
QY 336 -----TCCAAACGACACAGTCAGCCCGCGAGCCCTCCGCATCGTACTATGAACAG 383
Db 109 SerArgAlaArgSerThrAspThrPheAsn-----TyrAlaThr 121
QY 384 TATCACTCACTAAATGAATCTATTCTTGGATAGAATTTATACTGAGCGCATCTCGAT 443
Db 122 TyrHisThrLeuGluGluIleTyrAspPheLeuAspLeuValAlaGluAsnProHis 141
QY 444 ATGCTTACAAAATCCACATCGATCTCTCATTTGAGAGTACCCTCTATGTTTAAAG 503
Db 142 LeuValSerLysIleGlnIleGlyAsnThrTyrGluGlyArgProIleTyrValLeuLys 161
QY 504 GTTCTCTGAAAGAAACAAACAGCCAAATATGCAATATGATGATGATGCAATCCATGCC 563
Db 162 PheSer---ThrGlyGlySerLysArgProAlaIleTrpIleAspThrGlyIleHisSer 180
QY 564 AGAATGATGATCTCTCTCTCTTCTGCTTGTGGTTCATA----- 602
Db 181 ArgGluTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThrGlnAspTyr 200
QY 602 ----- 602
Db 201 GlyGlnAspAlaPheThrAlaIleLeuAspThrLeuAspIlePheLeuGluIleVal 220
QY 603 -----GGCCATAATCGAATGTGAGAGAAAGACCGT 632
Db 221 ThrAsnProAspGlyPheAlaPheThrHisSerThrAsnArgMetTrpArgLysThrArg 240
QY 633 TCTTCTTATCGAACAATCATTCATCGGACACGACCTGAATAGCAATTTGTCTCAAA 692
Db 241 SerHisThrAlaGlySerLeuCysIleGlyValAspProAsnArgAsnTrp---AspAla 259
QY 693 CACTGGTGTGAGNAGTGATCCAGTCTCTCATGCTCGGAAACCTACTGTGGACTTAT 752
Db 260 GlyPheGlyLeuSerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyLysPhe 279
QY 753 CCTGAGTCAGAACAGAGTGAAGCGAGTGGCTAGTTTCTTGAAGAAATATCAACACAG 812
Db 280 AlaAsnSerGluValGluValLysSerIleValAspPheValLysAspHisGlyAsn--- 298
QY 813 ATTAAGCATATCATCAGCATGCAATTCATCTCCAGCATATAGTGTGTTTCCATATTCCTAT 872
Db 299 IleLysAlaPheIleSerIleHisSerTyrSerGlnLeuMetTyrProTyrGlyTyr 318
QY 873 ACAGAAAGTAAACCAAGACCATGAGCACTGCTCTAGTACCCAGTGAAGCAGTTCGT 932
Db 319 LysThrGluProValProAspGlnAspGluLeuAspGlnLeuSerLysAlaAlaValThr 338
QY 933 GCTATTGACAAAACCTAGTAAAAATACCAAGGTATACATACATGCGCATGCGCTCAGAAACCTTA 992
Db 339 AlaLeu---AlaSerLeuTyrGlyThrLysPheAsnTyrGlySerIleIleLysAlaIle 357
QY 993 TACCTAGCTCTGAGGTGGGACGATTGGATCTATGATTGGGCATCAAAATATTCGTTT 1052
```

```
Db 358 TyrGlnAlaSerGlySerThrIleAspTrpThrTyrSerGlnGlyIleLysTyrSerPhe 377
QY 1053 ACA 1055
Db 378 Thr 378
RESULT 13
T33527
hypothetical protein T06A4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
R;Accession: T33527
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33527
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-528 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.1
A;Map position: 1
A;Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C;Superfamily: carboxypeptidase
Alignment Scores:
Pred. No.: 1,92e-30 Length: 528
Score: 437.50 Matches: 129
Percent Similarity: 45.28% Conservative: 63
Best Local Similarity: 30.42% Mismatches: 139
Query Match: 15.75% Indels: 93
Db: 2 Gaps: 16
US-09-980-881A-1 (1-1573) x T33527 (1-528)
QY 18 ATCAAGCTTTGCAGC---CTTGCACTCTGTGTACCATTTCTTCTCTCTGTGAGCAGCAT 74
Db 1 MetThrLysCysGlnIleLeuAlaSerLeuLeuThrTrpSerPheTrpValPheArgHis 20
QY 75 GTCTTCGCTTCCAGAGTGGC-----CAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTT 95
Db 21 ValLeuAlaThrThrAspGlyArgAlaPheArgLysLysSerSerAlaThrThrProPro 40
QY 96 -----CAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTT 134
Db 41 AsnGlnLysThrSerSerPheLysLeuIleArgIleAsnProGluThrGluGlySerVal 60
QY 135 CAAGTTCTACAGATCTT-----ACTACAACATATGAGATTGTTCTCTGCGAGCCG 185
Db 61 LysTyrLeuArgSerLeuTyrGluAspProSerProTyrGluLeuAspPheTrpGlnPro 80
QY 186 GTAACA-----GCTGACCTTATTGTGAAGAAAAACAAGTCCATTTT 230
Db 81 ProThrAsnIleGlyAlaIleValAspLeuThrValAlaProAlaAspAlaProArgPhe 100
QY 231 GTAATGCACTCTCATGTCGACAAATGAAAGCCCATTTAAATGTGAGCGCAATTCATGC 290
Db 101 Val-----LysAspLeuGluSerLysLys-----IleSerTyr 111
QY 291 AGTGTCTTCTGCGACAGCTGGAGATCTTATTCAACAGCAGATTTCACACGACAGTC 350
Db 112 IleValAlaValAsnAspLeuSerLysAlaIleGluAsnGluArgGlySerAspLysPhe 131
QY 351 AGCCCCGCGAGCTCCGCATCGTACTATGAACAGTATCACTCACTAAATAAATCTATTC 410
Db 132 TyrAsnProValAlaGlyPheAlaTyrAspLysTyrAsnSerLeuGluGluIleGlnThr 151
QY 411 TGGATGAAATTTAATACTGAGAGCATCTGATGATGCTTACAAAAATCCATTTGATCC 470
Db 152 GluMetLysArgLeuLysLysGluTyrProThrMetIleThrLeuIleAspIleGlyGln 171
```

```
QY 471 TCATTGAGAGTACCACCTCTATGTTTAAAGTTTCTGGAAGAACAAACA----- 524
Db 172 SerHisGluAenArgThrLeuLeuValMetLysIleThrGlyLysArgAenProLeuGly 191
QY 525 GCCAAAATGCATATGATGACTGTGGAATCCATCCAGAGAAATGATCTCTCCGTCT 584
Db 192 SerLysIleSerMetTrpIleAspAlaGlyIleHisAlaArgGluTrpIleAlaProAla 211
QY 585 TTCTGCTGTGGTTCATAGCCAT----- 608
Db 212 ThrAlaMet---TyrIleAlaHisGluLeuLeuGlyTyrGluAenAspAlaThrVal 230
QY 608 ----- 608
Db 231 AlaLysLeuMetAspHisIleAspPheTyrIleLeuProValMetAenProAspGlyTyr 250
QY 609 -----AATCGAATGTGGAGAAGAACCGTTCT----- 635
Db 251 GluTyrSerArgGluLysAenArgMetTrpArgLysAenArgSerProAlaLysCysAla 270
QY 636 -----TTCTATGCGAACAATCATTTGCATCGGAACAGACCTGAACTTGTCTCC 689
Db 271 ArgGlnThrPheSerThrValCysCysSerGlyValAspLeuAenArgAenPhe---Asp 289
QY 690 AAACACTGGTGTGGAGAGTGATCCAGTTCCTCATCTGCTCGGAACCTACTGTGGACTT 749
Db 290 TrpPheTrpAlaSerThrGlySerSerAspProCysHisAspThrTyrHisGlySer 309
QY 750 TATCTCTGAGTCAGAACAGAGTGAAGCGATGCTAGTTCTTGTGAGAAGAAATATCAAC 809
Db 310 AlaAlaPheSerGluProGluSerGlnAlaValArgAspPheLeuGluGlnAenThrPro 329
QY 810 CAGATTAAAGCATACATCAGCATGCTATCTATCTCCAGCATATAGTGTTCATATTC 869
Db 330 Glu-----AlaPheIleSerLeuHisSerTyrSerGlnMetTrpLeuIleProTyrGly 347
QY 870 TATACAGCA---AGTAAAGCAAGAC---CATGAGGAATCTCTCTAGTAGGACAGGAA 923
Db 348 HisArgLysGlnSerTyrProGlnAspTyrHisThrGlyLeuArgProLeuAlaLeuArg 367
QY 924 GCAGTTCTGCTATTGACAAACTAGTAAATAACAGTATACATACATGCGCATCTCA 983
Db 368 AlaThrLysAlaLeuTyrGluLeu---TyrGlyThrLysTyrGlnValGlyThrGlyAla 386
QY 984 GAAACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGGATCTAT---GATTTGGGCATC 1040
Db 387 AspLeuMetTyrGluAlaSerGlyGlySerHisAspTrpAlaLysGlyGlnLeuLysVal 406
QY 1041 AATATTCTGTTT 1052
Db 407 ProTyrAlaTyr 410
```

RESULT 14

T33526

hypothetical protein T06A4.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33526

R:Wu, X.; Tin-Wollam, A.; Ozeresky, P.; Wilson, R.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of C. elegans cosmid T06A4.

A:Reference number: Z21364

A:Accession: T33526

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-667 <WUX>

A:Cross-references: EMBL:AF098994; PIDN:AA6G7474.1; GSPDB:GN00019; CESP:T06A4.3

A:Experimental source: strain Bristol N2; clone T06A4

C:Genetics:

A:Gene: CESP:T06A4.3

A:Map position: 1

A:Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 582/3

```
Alignment Scores:
Pred. No.: 3,73e-27 Length: 667
Score: 400.50 Matches: 131
Percent Similarity: 40.70% Conservative: 66
Best Local Similarity: 27.07% Mismatches: 148
Query Match: 14.42% Indels: 139
DB: 2 Gaps: 16
```

US-09-980-881A-1 (1-1573) x T33526 (1-667)

```
QY 51 ATTGTTCTCTCTGTCGAGCAGCATGTC---TTCCGGTTCGAGAGTGCCAGTCTTAGCT 107
Db 8 LeuIlePheLeuCysPheSerAlaValSerLeuSerThrGlnSerPheAlaValTyrArg 27
QY 108 GCTCTTCTCTAGAACCTCTAGGCAAGTTCAAAGTTCTACAGAAATCTT-----ACTACA 158
Db 28 LeuLeuProLysSerGlnThrAspPheGlnAlaIleGlnArgLeuTyrLysAsnAlaThr 47
QY 159 ACATATGAGATTGTTCTCTGCGACCGGTAAACAGCT-----GACCTTATT 203
Db 48 AspHisAspLeuAenPheTrpLysThrGlyLysAspLysHisGlyPheTyrAspValMet 67
QY 204 GTCAAGAAAAACAAGTCCATTTTGTAAATGCATCTCATGTCGACAAATGTGAAAGCC 263
Db 68 ValAspMetLysAenSerTyrPheLeuAsp----- 78
QY 264 CATTTAAATGTGCGCGAATTCCATGCAAGTGTCTGCTGCGACAGCTGGAAGATCTTATT 323
Db 79 PheLeuGlnValAsnAspIleSerTyrIleLysThrIleAspAspValGluGlyLeuIle 98
QY 324 CAACAGCAGATTTCACACGACAGTCAGC-----CCCCGA----- 359
Db 99 ThrLysHisGluAenAenLysThrLeuSerAsnMetPheProArgLeuTyrAspAspSer 118
QY 360 GCCTCCGCGATCGTAC---TATGAACAGTATCACTCACTAAATGAATCTATTCTTGATA 416
Db 119 SerSerAlaHisTyrAspPheHisThrTyrGlySerTyrGlnArgMetThrAspTrpMet 138
QY 417 GAATTATTAACCTGAGAGGCATCTGATATGCTTACAAAATCCACATTTGGATCTCTATT 476
Db 139 LysGlnLeuValValLysTyrProLysMetValGlnTyrIleSerIleGlyLysThr 158
QY 477 GAGAAGTACCCTCTATGTTTAAAGTTTCTGAAAAGAACAAACAGACCAAAATGCC 536
Db 159 GluGlyArgAsnIleAspGlyValGluIleGlyLysAspSerArgThr---LysLysIle 177
QY 537 ATATGATTGACTGTGGAATCCATGCCAGAGATGGAATCTCTCTGCTTCTCTGCTGG 596
Db 178 PheTrpIleAspGlyIleHisAlaArgGluTrpAlaAlaProHisThrAlaLeuPhe 197
QY 597 TTCATA----- 602
Db 198 PheIleHisGlnValCysAspProAlaProSerGluLysSerValPheArgLeuValPro 217
QY 602 ----- 602
Db 218 LeuGlyLeuHisLysCysTrpGlnArgGlnLysLysGlyLysLeuAspAsnLysTrpLeu 237
QY 602 ----- 602
Db 238 ThrSerArgAlaAsnGluProGlyIleLysLeuLeuAenGluIleThrPheValVal 257
QY 603 -----GCCATAAT 611
Db 258 ValProCysLeuAenProAspGlyTyrGluPheThrArgSerSerThrAenProHisVal 277
QY 612 CGAATGTGGAAGAAAGACCGTTCT-----TTCTATGCGAAC 647
Db 278 ArgLeuTrpArgLysAsnArgSerLysMetGlnCysArgLysAspIleTrpGlyLysArgAen 297
QY 648 AATCATTTGCATCGGAACAGACCTGAAATAGCAACTTTGCTCTCCAAACACCTGGTGTGGAA 707
```

Db 298 ArgCysCysArgGlyValAspLeuAsnArgAsnPhe---AspPheHisPheArgGluSer 316
QY 708 GGTGCATCCAGTTCCTCATCTCGGAACCTACTGTGGACTTTATCCTGAGTCAGAACCA 767
Db 317 GlyThrSerAspProCysSerGluIleTyrGlnGlyProSerProPheSerGluPro 336
QY 768 GAAGTGAAGGCAGTGGCTAGTCTTCTG-----AGAAGAATATCAACAGATTAAAGCA 821
Db 337 GluAlaLysAlaValArgAspAlaLeuSerGlnArgTyrLysGlyArgThrAspAla 356
QY 822 TACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTATACACGA--- 878
Db 357 TyrIleThrLeuHisThrTyrSerGlnIleTyrPheHisProTyrGlyHisLysLysAsp 376
QY 879 AGTAAAGCAAGACCATGAGGAAGTCTCTCTAGTACCCAGTGAAGCAGTTCGTATT 938
Db 377 AlaTyrProGlyAspLysLeuTyrGluValGlyLysAlaAlaGlnAlaLeu 396
QY 939 GACAAACTAGTAAATAATACAGGTATACATGCGCCATCGCTCAGAAACCTTATACCTA 998
Db 397 ---LysArgValTyrGlyThrLysTyrValValGlySerGlyAlaAspThrLeuTyrPro 415
QY 999 GCTCTGGAGTGGGACGATGGATC---TATGATTGGGCATCAATAATTCGTTTACA 1055
Db 416 AlaSerGlyGlySerGluAspTyrAlaLysHisGluAlaLysValLys--PheValTyrL 435
QY 1056 TCAAA----- 1060
Db 435 euLeuGluLeuArgProAspGluLysAsnTrpAspGlyPheIleLeuAspGluLysGluL 455
QY 1061 -----CCACCTGTAGAGAGCTTTGGCCGCTGCTCTAAATAAGCTTGGCATGTCA 1115
Db 455 euIleProThrAlaArgGluThrTrpGluGlyValArgValAlaGluAlaValLeuA 475
QY 1116 GGAATGTT 1123
Db 475 spArgile 477

RESULT 15
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20507; T24289
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19284
A;Accession: T20507
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-323 <WIL>
A;Cross-references: EMBL:Z78411; PIDN: CAB01647.1; GSPDB: GN00023; CBSP: F02D8.4
A;Experimental source: clone F02D8
R;Wild, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19869
A;Accession: T24289
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-323 <W12>
A;Cross-references: EMBL:Z78413; PIDN: CAB01666.1; GSPDB: GN00023; CBSP: F02D8.4
A;Experimental source: clone T01C3
C;Genetics:
A;Gene: CBSP: F02D8.4
A;Map position: 5
A;Introns: 33/3; 67/3; 201/3; 228/1; 257/2
C;Superfamily: carboxypeptidase

Alignment Scores:
Pred. No.: 6.97e-27 Length: 323
Score: 397.00 Matches: 88
Percent Similarity: 53.01% Conservative: 53
Best Local Similarity: 33.08% Mismatches: 81

Query Match: 14.30% Indels: 44
DB: 2 Gaps: 7
US-09-980-881A-1 (1-1573) x T20507 (1-323)
QY 411 TGGATAGAATTTATATACTGAGAGGCATCCTGATATGCTTACAAAAATCCATTTGGATCC 470
Db 3 PheLeuAsnSerLeuAlaGlnGlnTyrProAsnAspValLysLeuGlnAsnIleGlyAsn 22
QY 471 TCATTGGAGATGACCCACTCTATGTTTAAAGTTTCTGGAAAAAGAACACAGCCAAA 530
Db 23 SerTyrGluGlyArgSerIleThrAlaValArgIleAla---AspAspGlySerSerLys 41
QY 531 AATGCCATTCGATTCACCTGTCGATCCATCCAGAGAAATGGATCTCTCTGCTTCTGCG 590
Db 42 ProIleValTrpIleAspAlaGlyIleHisAlaArgLysIleSerTyrAsnValAla 61
QY 591 TTGTGGTTCATA----- 602
Db 62 LeuTyrLeuIleTyrThrIleValSerGlnProAlaTyrArgAsnLeuLeuAspSerVal 81
QY 603 -----GCCATAAT 611
Db 82 GlnLeuValValProAsnThrAsnProAspGlyTyrGluTyrSerArgThrAsnAsp 101
QY 612 CGAATCTGGAGAAAGAACCGTCTTCTTATCGCAACAATCATTTGCATCGGAACAGACCTG 671
Db 102 ArgMetTrpArgLysThrArgSerArgPheThrAsnSerArgCysAlaGlyAlaAspAla 121
QY 672 AATAGCAACTTTCTCTCAAACTGCTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCG 731
Db 122 AsnArgAsnTyr---ProPheTyrTrpGlyThrGlnGlyValSerHisSerGlnCysSer 140
QY 732 GAAACCTACTGTGAGCTTTATCTGAGTCAGAACCAAGAGTGAAGCGAGTGGCTAGTTTC 791
Db 141 GluIlePheCysGlySerArgProGlnSerGluProGluValLeuAlaLeuThrAsnAla 160
QY 792 TTGAGAGAATATCAACCAAGATTAAGCATACATCAGCATGCATTCATCTACTCCAGCAT 851
Db 161 IleIleArgAspGluGluArgIleLysGlyTyrIleAlaLeuHisSerTyrGlyGlnGlu 180
QY 852 ATAGTGTTCATATTCCTATACA---CGAAGTAAAGCAAGAACCATCAGGAACTGTCT 908
Db 181 IleLeuTyrProTrpGlyHisThrGlnArgThrTyrProThrAspValGlnAspLeuIle 200
QY 909 CTAGTAGCC-----AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959
Db 201 GlnValGlyArgAlaMetAlaSerAlaIleArgAlaValAsn-----AsnThr 216
QY 960 AGGTATACATGCGCCATGCTGCTCAGAAACCTTATACCTAGTCTCTGGAGTGGGACGAT 1019
Db 217 AspTyrThrValValAsnSerGlyAspGlyLeuTyrProAlaAlaGlyAlaSerAspAsp 236
QY 1020 TGGATCTATGATTGGGCATCAATATTCGTTTACATCAAC---CCACCTGTAGAGAAG 1076
Db 237 TrpAlaLysSerArgGlyIleLysTyrSerTyrThrIleGluLeuSerProIleAspAsp 256
QY 1077 CTTTGGCCGCTGCTCTA 1094
Db 257 PheThrGlyPheSerLeu 262

Search completed: January 1, 2004, 19:00:31
Job time : 57.8315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:41:10 ; Search time 16.6016 Seconds
(without alignments)
8911.561 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgtggatg9.....aaaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Zgapop 6.0	Zgapext 7.0
Delop 6.0	Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DBV=xlh
-O=/cgn2_1/USPTO.spool/US0980881/runat_31122003_135748_24941/app_query.fasta_1.2958
-DB=SwissProt_41 -QWTF=fastan -SUFFIX=r5p -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0980881 @CGN 1.1.34 @runat_31122003_135748_24941 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	695.5	25.0	415	1 CBPB RAT	P19223 rattus norv
2	655	23.6	417	1 CBPC MOUSE	P15089 mus musculu
3	652.5	23.5	417	1 CBPB HUMAN	P15086 homo sapien
4	647.5	23.3	417	1 CBPC HUMAN	P15088 homo sapien
5	640.5	23.1	416	1 CBPB CANFA	P55261 canis famil
6	629.5	22.7	401	1 CBPB PIG	P09955 sus scrofa
7	589.5	21.2	306	1 CBPB BOVIN	P00732 bos taurus
8	567.5	20.4	309	1 CBPC RAT	P21961 rattus norv
9	553	19.9	417	1 CBP2 RAT	P19222 rattus norv
10	536	19.3	419	1 CBPA BOVIN	P00730 bos taurus
11	530	19.1	419	1 CBP1 RAT	P00731 rattus norv
12	511	18.4	417	1 CPB2 HUMAN	P48052 homo sapien
13	489.5	17.6	419	1 CBP1 HUMAN	P15085 homo sapien
14	483	17.4	421	1 CBP4 HUMAN	O9ui42 homo sapien
15	409	14.7	304	1 CBP2 SIMVI	P42788 simulium vi
16	386	13.9	433	1 CBPA ANOCA	O02350 anopheles g
17	369.5	13.3	303	1 CBPB ASTFL	P04059 astacus flu
18	299.5	10.8	430	1 YHT2 YEAST	P38836 saccharomyc

19	277.5	10.0	424	1	CBPT THEVU	P29068 thermoactin
20	277	10.0	451	1	CBPS STRGR	P18143 streptomyc
21	252.5	9.1	434	1	CBPS STRCP	P39041 streptomyc
22	118.5	4.3	707	1	ORC1 SCHPO	P54789 schizosacch
23	114.5	4.1	1380	1	CBPD HUMAN	O75976 homo sapien
24	112.5	4.1	1377	1	CBPD MOUSE	O89001 mus musculu
25	112	4.0	1368	1	RRPO PYMR	P17965 potatoco viru
26	108.5	3.9	1255	1	PER2 HUMAN	O15055 homo sapien
27	103.5	3.7	722	1	CPXM MOUSE	O92100 mus musculu
28	102	3.7	734	1	CPXM HUMAN	O96m3 homo sapien
29	100	3.6	376	1	YQGT BACSU	P54497 bacillus su
30	99.5	3.6	445	1	TPH1 CHICK	P70080 gallus gall
31	98.5	3.5	488	1	TY3H ANGAN	O42091 anguilla an
32	98.5	3.5	1160	1	C8CA BACTP	O45706 bacillus th
33	97	3.5	663	1	TAZI SCHPO	P79005 schizosacch
34	95	3.4	444	1	TPH1 HUMAN	P17752 homo sapien
35	95	3.4	1793	1	YCF1 LOTJA	O9bbn6 lotus japon
36	94.5	3.4	482	1	YPT1 CAEEL	P41879 caenorhabdi
37	94.5	3.4	744	1	YG13 YEAST	P53202 saccharomyc
38	94.5	3.4	2185	1	PYRI DICDI	P20054 dictyosteli
39	94	3.4	435	1	YNJE ECOLI	P78067 escherichia
40	93	3.3	853	1	PHS1 DICDI	Q00766 dictyosteli
41	92	3.3	525	1	PUR9 BUCAI	P57143 b bifunctio
42	92	3.3	835	1	VP3 ROTSI	P57326 simian 11 r
43	91.5	3.3	465	1	STHA ECO57	Q8x727 escherichia
44	91.5	3.3	465	1	STHA ECOLI	P27306 escherichia
45	91.5	3.3	982	1	ENV_SPV3L	P27399 simian foam

ALIGNMENTS

RESULT 1

ID	CBPB RAT	STANDARD	PRT	415 AA
AC	P19223			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carboxypeptidase B precursor (EC 3.4.17.2)			
GN	CPB			
OS	Rattus norvegicus (Rat)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=101116;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=89034324; PubMed=3182872;			
RA	Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;			
RT	"Structural characterization of the rat carboxypeptidase A1 and B			
RT	genes. Comparative analysis of the rat carboxypeptidase gene			
RT	family."			
RL	J. Biol. Chem. 263:17837-17845(1988).			
CC	-I- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =			
CC	peptide + L-lysine (or L-arginine).			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M23959; AAA40872.1;			
DR	EMBL; M23947; AAA40872.1; JOINED.			
DR	EMBL; M23950; AAA40872.1; JOINED.			
DR	EMBL; M23952; AAA40872.1; JOINED.			
DR	EMBL; M23953; AAA40872.1; JOINED.			
DR	EMBL; M23954; AAA40872.1; JOINED.			
DR	PIR; A32129; A32129.			
DR	HSSP; P09955; INSA.			
DR	MEROPS; M14.003; -.			

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; J05118; AAA37369.1; -.
CC PIR; A34487; A34487.
CC DR HSSP; P09955; INSA.
CC DR MEROPS; M14.010; -.
CC DR MGD; MGI:88479; Cpa3.
CC DR InterPro; IPR003146; Propep_M14.
CC DR InterPro; IPR00834; Zn_carboPept.
CC DR Pfam; PF02244; Propep_M14; 1.
CC DR Pfam; PF02246; Zn_carboPept; 1.
CC DR SMART; SM00631; Zn_dept; 1.
CC DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
CC KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
CC FT SIGNAL 1 15
CC FT PROPEP 16 109
CC FT CHAIN 110 417
CC FT METAL 176 176
CC FT METAL 179 179
CC FT METAL 304 304
CC FT ACT_SITE 378 378
CC FT DISULFID 173 186
CC FT DISULFID 245 268
CC SQ SEQUENCE 417 AA; 48790 MW; A2B300A068D1BA6D CRC64;

Alignment Scores:
Pred. No.: 2,54e-50 Length: 417
Score: 655.00 Matches: 150
Percent Similarity: 51.67% Conservative: 66
Best Local Similarity: 35.89% Mismatches: 138
Query Match: 23.59% Indels: 65
DB: Gaps: 9

US-09-980-881A-1 (1-1573) x CBPB_MOUSE (1-417)
QY 33 CTTCGAGTCCTTACCATGTTCTCTCTGTGAGCAGCATGCTTCGCGTTCCAGAGT 92
DB 7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis-----PheAspArg 23
QY 93 GGCACAGTCTAGCTGCTCTCTCTAGAACCTCTAGGCAAGTTCAGTCTTACAGATCTT 152
DB 24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43
QY 153 ACTACACATATGAGATTGTTCTCTGCGACCGGTAAACAGCTGACCTTATTGTGAAGAA 212
DB 44 ThrGlnSerIleGluLeuAspPheTyrTyrProAspAlaIleHisAspIleAlaValAsn 63
QY 213 AAACAAGTCCATTTTGTGTAATGTCATCTGTGTCGACAAATGTGAAGCCCAATTTAAAT 272
DB 64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83
QY 273 GTGAGCGGAATTCAGTCAGTCTGCTGCGACAGTGGAAGATCTTATTCAACAGCAG 332
DB 84 GlnHisIleHisIleHisTyrGluIleLeuIleHisAspLeuGlnGluIleGluLysGln 103
QY 333 ATTTCC---AACACACAGTCAGCCCGCCGCGCTCGCATCTGACTGACAGATGATCAC 389
DB 104 PheAspValLysAspGluIleAlaGlyArgHisSer-----TyrAlaLysTyrAsn 120
QY 390 TCACATAATGAATCTATTCTTGGATAGAAATTTATTAACGTAGGAGCATCTCTGATGCTT 449
DB 121 AspTyrAspLysIleValSerTyrThrGluLysMetLeuLysHisProGluMetVal 140
QY 450 ACAAAATCCATGTCATCTCATTTGAGAAATGACCATCTATGTTTAAAGGTTTCT 509
DB 141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159
QY 510 GGAAGAAGAACACACGCAAAATGCATATGATGATGATGATGATGATGATGATGATGATG 569
DB 160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

```

```

QY 570 TGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 608
DB 180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199
QY 609 AAT----- 611
DB 200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219
QY 612 -----CGAATGTGGAGAAGAACCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 638
DB 220 ValAspGlyTyrIleTyrSerTyrThrGlnAspArgMetTyrArgLysAsnArgSerArg 239
QY 639 TATGCCAAACAATCATTCGATCGAAGACAGACCTGCAATAGCACTTGTCTCTCAACACTGG 698
DB 240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258
QY 699 TGTGAGGAAGGTGCATCCAGTTCCTCATGTCGGAACCTACTGTGAGCTTTATCTCTGAG 758
DB 259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278
QY 759 TCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTCGAGAAGAAATATCAACAGATTAAA 818
DB 279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
QY 819 GCATACATCAGCATGCAATTCATCTACTCCAGCATATAGTGTTCATATTCCTATACACGA 878
DB 299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuLeuProTyrGlyTyrThrPhe 318
QY 879 AGTAAAGCAAGACCATGAGGAACGTCTCTAGTAGCAGCAGTCAAGCAGTTCGTGCTATT 938
DB 319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338
QY 939 GACAAACATAGTAAATAATACACAGTATACACATGGCCATGGCTCAGAACCTTATACCTA 998
DB 339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357
QY 999 GCTCTCGAGGTGGGAGCAGTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1054
DB 358 ThrSerGlySerSerLeuAspTyrValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377
QY 1055 -----ATCAAAACCCACC 1066
DB 378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397
QY 1067 TGTAGAGAAGAGCTTTTGCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1120
DB 398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415

RESULT 3
CBPB_HUMAN
ID CBPB_HUMAN STANDARD; PRT; 417 AA.
AC P15086; O60834; Q96B08;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
DE (PASP).
DE CPB1 OR CPB OR PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC TISSUE=Pancreas;
RX MEDLINE=92129345; PubMed=1370825;
RA Yamanoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA French C.K.;
RT "Isolation of a cDNA encoding a human serum marker for acute
RT pancreatitis. Identification of pancreas-specific protein as
RT pancreatic procarboxypeptidase B."
RL J. Biol. Chem. 267:2575-2581(1992).
RN [2]

```


Db 237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProenArgAsnPhne---Asp 255
Qy 690 AAACACTGGTGTGAGGAGGTGCATCCAGTTCCTCATCGCGAAACCTACTGTGGACTT 749
Db 256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrcysGlyPro 275
Qy 750 TATCCTGAGTCAGACAGAGGAGTGAAGCGAGTGGCTAGTCTTCTTGAGAGAANAATATCAAC 809
Db 276 AlaAlaGluSerGluTyrsGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
Qy 810 CAGATTAAGCATACATCAGCATCATTCTACTATCCAGCATAGTAGTGTTCATATCC 869
Db 296 SerIleLysAlaTyrluThrIleHisSerTyrsGlnMetMetIleTyrlProTyrs 315
Qy 870 TATACACGAAGTAAACCAAGACCATCAGGAACTGCTCTCTAGTAGCCAGTGAAGCAGTT 929
Db 316 TyrAlaTyrlLysLeuGlyGluAsnAsnAlaGluLeuAlaLeuAlaLysAlaThrVal 335
Qy 930 CGTGTATTGACAAACTAGTAAATACAGGATATACATGCGCCATGCTCAGAAACC 989
Db 336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrlThrTyrlGlyProGlyAlaThrThr 354
Qy 990 TTATACCTAGCTCTGAGGTGGGAGCATGATGATCATGATTTGGGCATCAATATTCG 1049
Db 355 IleTyrlProAlaAlaGlyGlySerAspTrpAlaTyrlAspGlnGlyIleA-gTyrlSer 374
Qy 1050 TTATACATC----- 1057
Db 375 PheThrPheGluLeuArgAspThrGlyArgTyrlGlyPheLeuLeuProGluSerGlnIle 394
Qy 1058 AAACCCACCTGTGAGAGCTTTGGCGTCTCTCTAAATAGCTGGCATGTCATPAGG 1117
Db 395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrlValAlaSerTyrlValLeuGlu 414
Qy 1118 AATGTT 1123
Db 415 HisLeu 416

RESULT 4
CBPC_HUMAN STANDARD; PRT; 417 AA.
AC P15088; Q96E94;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=90083291; PubMed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mast cells;
RX MEDLINE=92105393; PubMed=1729276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A.";
RL J. Clin. Invest. 89:273-282(1992).
RN [3]
RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapletoen M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heiton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 110-417 FROM N.A.
RX MEDLINE=92333185; PubMed=1629626;
RA Natsuki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuki M.,
RA Wintroub B.U., Rutter W.J., Goldstein S.M.;
RT "Human skin mast cell carboxypeptidase: functional characterization,
RT cDNA cloning, and genealogy.";
RL J. Invest. Dermatol. 99:138-145(1992).
RN [5]
RP SEQUENCE OF 110-137.
RX MEDLINE=892114692; PubMed=2708524;
RA Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT "Human mast cell carboxypeptidase. Purification and
RT characterization.";
RL J. Clin. Invest. 83:1630-1636(1989).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -I- SUBCELLULAR LOCATION: Secretory granules.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; M27717; AAA35652.1; -
DR ENBL; M73720; AAA59568.1; -
DR ENBL; M73716; AAA59568.1; JOINED.
DR ENBL; M73717; AAA59568.1; JOINED.
DR ENBL; M73718; AAA59568.1; JOINED.
DR ENBL; M73719; AAA59568.1; JOINED.
DR ENBL; BC012613; AAH12613.1; -
DR ENBL; S40234; AAB22578.2; ALT_SEQ.
DR PIR; A43929; A43929.
DR HSP; P09955; INSA.
DR MEROPS; M14.010; -
DR Genew; HGNC:2298; CPA3.
DR MIM; 114851; -
DR GO; GO:0030141; C:secretory granule; NAS.
DR GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; ZnCarbOpept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; ZnCarbOpept; 1.
DR SMART; SM00631; zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.

structural basis of its inactivity.";
 [4]
 STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
 MEDLINE=91027767; PubMed=2223783;
 Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
 "Sequence-specific 1H NMR assignments and determination of the
 secondary structure for the activation domain isolated from
 pancreatic procarboxypeptidase B.";
 Biochemistry 29:7515-7522(1990).
 [5]
 STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
 MEDLINE=91114693; PubMed=1989879;
 Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
 "The NMR structure of the activation domain isolated from porcine
 procarboxypeptidase B.";
 EMBO J. 10:11-15(1991).
 [6]
 STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
 MEDLINE=93044373; PubMed=1422143;
 Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
 Huber R., Wuethrich K.;
 "Comparison of the NMR solution structure with the X-ray crystal
 structure of the activation domain from procarboxypeptidase B.";
 J. Biomol. NMR 2:1-10(1992).
 -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =
 peptide + L-lysine (or L-arginine).
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
 -1- DATABASE: NAME=worthington enzyme manual;
 PDB; 1PBA; 31-OCT-93.
 PDB; 1NSA; 24-DEC-97.
 MEOPS; M14.003;
 InterPro; IPR003146; Propep M14.
 InterPro; IPR00834; Zn carbOpept.
 Pfam; PF02244; Propep M14; 1.
 Pfam; PF00246; Zn carbOpept; 1.
 PRINTS; PR00765; CRBOXYPTASEA.
 SMART; SM00631; Zn.pept. 1.
 PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
 3D-structure.
 PROPEP 1 95 ACTIVATION PEPTIDE.
 CHAIN 96 401 CARBOXYPEPTIDASE B.
 DISULFID 158 171
 DISULFID 230 253
 DISULFID 244 258
 METAL 161 161 ZINC.
 METAL 164 164 ZINC.
 METAL 289 289 ZINC.
 ACT_SITE 341 341
 ACT_SITE 363 363 NUCLEOPHILE.
 TURN 8 9
 STRAND 11 17
 HELIX 20 32
 STRAND 36 39
 STRAND 43 45
 STRAND 50 56
 STRAND 58 60
 HELIX 61 69
 TURN 70 72
 STRAND 74 79
 HELIX 82 87
 TURN 88 88
 TURN 101 102
 HELIX 107 120
 TURN 122 124
 TURN 125 132
 STRAND 134 135
 STRAND 138 144
 TURN 149 150
 STRAND 153 157

TURN 162 163
 HELIX 165 181
 TURN 182 184
 HELIX 186 194
 STRAND 196 200
 HELIX 205 213
 TURN 214 214
 TURN 216 217
 STRAND 224 224
 TURN 226 227
 STRAND 231 231
 HELIX 235 237
 TURN 243 244
 STRAND 249 249
 TURN 252 253
 TURN 255 256
 STRAND 257 257
 TURN 262 263
 HELIX 266 277
 TURN 278 281
 STRAND 282 289
 STRAND 294 297
 TURN 307 308
 HELIX 309 327
 STRAND 332 335
 HELIX 336 339
 HELIX 347 353
 TURN 354 355
 STRAND 359 363
 TURN 371 372
 HELIX 376 378
 HELIX 379 398
 TURN 399 401
 SQ SEQUENCE 401 AA; 45713 MW; 53129AF159A26348 CRC64;
 Alignment Scores:
 Pred. No.: 4.67e-48 Length: 401
 Score: 629.50 Matches: 134
 Percent Similarity: 54.85% Conservative: 64
 Best Local Similarity: 37.12% Mismatches: 120
 Query Match: 22.67% Indels: 43
 DB: 1 Gaps: 6
 US-09-980-881A-1 (1-1573) x CBPB_PIG (1-401)
 QY 84 TTCAGAGTGGCCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
 D 7 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerGluLeu 26
 QY 144 CAGAACTCTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGACCTTATT 203
 D 27 HisGluLeuAlaSerThrArgGlnIleAspPheTrpIysProAspSerValThrGlnIle 46
 QY 204 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGTATGTCGCAATGTGAAGACC 263
 D 47 LysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGluAsp 66
 QY 264 CATTAAATGTGACGGGAATTCATGCATGCGAGTGTCTTCTGCGCAGACGCTGGAAGATCTTATT 323
 D 67 PheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerValLeu 86
 QY 324 CAACAGCAGATTTCACACGACACAGTCAGCCCGCCGCTCCGATCGTACTACTGAACAG 383
 D 87 GluAlaGlnPheAspSerArgCys-----ArgThrThrGlyHisSerThrGluLys 103
 QY 384 TATCACTCACTAAATGAATCTATTCTTGATAGAAATTTATACTAGAGCGCATCTGAT 443
 D 104 TyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValThrSerLysAsnProAsp 123
 QY 444 ATGCTTACAAAAATCCACATGTCATTTGAGAGAGTACCACCTCTATGTTTAAAG 503
 D 124 LeuIleSerArgSerAlaIleGlyThrThrPheAspGlyAspAsnIleTyrLeuLys 143

```
QY 504 GTTCTCGAAGAACAAACAGCCCAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
Db 144 Val---GlyLysProGlySerAsnLysProAlaIlePheMetAspCysGlyPheHisAla 162
QY 564 AGAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 602
Db 163 ArgGluTrpIleSerGlnAlaPheCysGlnTrpPheValArgAspAlaValArgThrTyr 182
QY 602 ----- 602
Db 183 GlyTyrGluAlaHisMetThrGluPheLeuAspAsnLeuAspPheTyrValLeuProVal 202
QY 603 -----GCCAT-----NATCGATGTGGAGAAACCGT 632
Db 203 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThrArg 222
QY 633 TCTTCTATCGAACAATCTGTCATCGAAGACCTGAATAGCACTTTGTCTCCAAA 692
Db 223 SerThrAsnAlaGlySerSerCysThrGlyThrAspProAsnArgAsnPhe---AsnAla 241
QY 693 CACTGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGCACTTAT 752
Db 242 GlyTyrCysThrValGlyAlaSerValAsnProCysAsnGluThrTyrCysGlySerAla 261
QY 753 CTGAGTCAGAACAGAGTGAAGCAGTGTAGTCTTCTTGAGAGAAATATCAACCAG 812
Db 262 AlaGluSerGluLysGluThrLysAlaLeuAspPheIleArgAsnAsnLeuSerSer 281
QY 813 ATTAAGCATACATCAGCATGCTTCTACTCCAGCATATAGTGTTCCTATTCCTAT 872
Db 282 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 301
QY 873 ACAGAAAGTAAAGCAAGACCATGAGGAATGTCTCTAGTCAGTGAAGCAGTTCGT 932
Db 302 AspTyrLysLeuProGluAsnAspAlaGluLeuAsnSerLeuAlaLysGlyAlaValLys 321
QY 933 GCTATTGACAAACTAGTAAATACACAGTATACATGCGCATGCGTCCAGAACCTTA 992
Db 322 GluLeu---AlaSerLeuTyrGlyThrSerTyrSerTyrGlyProGlySerThrThrIle 340
QY 993 TACTAGCTCTGAGGTGGGACGATGATGATGATGATGATGATGATGATGATGATGAT 1052
Db 341 TyrProAlaAlaGlyLysSerAspSerTrpAlaTyrAsnGlnGlyIleLysTyrSerPhe 360
QY 1053 ACA 1055
Db 361 Thr 361
RESULT 7
CBPB_BOVIN STANDARD; PRT; 306 AA.
AC P00732;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN CPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
[2]
RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RX MEDLINE=74260705; PubMed=4833744;
RA Schmidt J.J., Hirs C.H.W.;
RT "Primary structure of bovine carboxypeptidase B. Inferences from the
```

```
RT locations of the half-cystines and identification of the active site
RT arginine.";
RL J. Biol. Chem. 249:3756-3764(1974).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX MEDLINE=76265065; PubMed=957425;
RA Schmid M.F., Harriott J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RL J. Mol. Biol. 103:175-190(1976).
RN [4]
RP ACTIVE SITE.
RX MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
carboxypeptidase B.";
RL J. Biol. Chem. 244:5246-5253(1969).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RT "Identification of a glutamic acid at the active center of bovine
carboxypeptidase B.";
RL J. Biol. Chem. 247:7864-7869(1972).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =
peptide + L-lysine (or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A93797; CPBOB.
DR PDB; 1CPB; 30-SEP-83.
DR MEROPS; M14.003; -.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT DISULFID 149 163
FT METAL 66 66 ZINC.
FT METAL 69 69 ZINC.
FT METAL 194 194 ZINC.
FT ACT_SITE 246 246
FT ACT_SITE 268 268 NUCLEOPHILE.
SQ SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;
Alignment Scores:
Pred. No.: 1,61e-44 Length: 306
Score: 589.50 Matches: 121
Percent Similarity: 56.58% Conservative: 51
Best Local Similarity: 39.80% Mismatches: 75
Query Match: 21.23% Indels: 58
DB: 1 Gaps: 6
US-09-980-881A-1 (1-1573) x CBPB_BOVIN (1-306)
QY 375 TATGAACAGTATCAGTCACTAAATGAATCTATCTTTGGATAGAAATTTATACTGAGAGG 434
Db 6 TyrGluLysTyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValAlaSerGlu 25
QY 435 CATCTGATATGCTTACAAAATCCACATTTGGATTCCTCATTTGAGAAGTACCCACTTAT 494
Db 26 AsnProAspLeuIleSerArgSerAlaIleGlyThrThrPheLeuGlyAsnThrIleTyr 45
QY 495 GTTTTAAAGTTTCTGGAAGAGNACAAACAGCCAAAATGCCATATGATGATGATGATGGA 554
Db 46 LeuLeuLysVal---GlyLysProGlySerAsnLysProAlaValPheMetAspCysGly 64
QY 555 ATCCATGACAGAGATGATCTCTCTGCTTCTCTGCTTCTGCTTCTGCTTCTGCTTCTATA 602
Db 65 PheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaVal 84
QY 602 ----- 602
```


Db 85 ArgThrTyrGlyArgGluIleHisMetThrGluPheLeuAspLysLeuAspPheTyrVal 104
QY 603 -----GGCCAT-----AATCGAATGCGAGA 623
Db 105 LeuProValValAsnIleAspGlyTyrIleTyrThrTrpThrAsnArgMetTrpArg 124
QY 624 AAGAACCGTTCTTCTATGCGAACAAATCATTTGCGAAGACAGACCTGGAATAGCAACTTT 683
Db 125 LysThrArgSerThrArgAlaGlySerCysThrGlyThrAspLeuAsnArgAsnPhe 144
QY 684 GTCTCCAAACACTGGTGTAGGAAGAGTGCATCCAGTCTCCATGCTCGGAACCTACTGT 743
Db 145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163
QY 744 GGACTTTATCTGAGTCAGACACAGAGTGAAGCAGTGGCTAGTCTTCTGAGAAGAAAT 803
Db 164 GlySerAlaAlaGluSerGluLysGluSerLysAlaValAlaAspPheIleArgHis 183
QY 804 ATCAACACAGATTAAGCATATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTC 863
Db 184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrPro 203
QY 864 TATTCCTATACAGAAAGTAAAGCAAGACATGAGGAAGTCTCTAGTAGGAGAGTAA 923
Db 204 TyrSerTyrAspTyrLysLeuProLysAsnAsnValGluLeuAsnThrLeuAlaLysGly 223
QY 924 GCAGTTCGTCTATTGACAAAACCTAGTAAATACACAGTATACACATGCGCATGCTCA 983
Db 224 AlaValLysLysLeu---AlaSerLeuHisGlyThrThrTyrSerTyrGlyProGlyAla 242
QY 984 GAAACCTTATACCTAGTCTCTGGAGTGGGACGATGATGATGATGATGATGATGATG 1043
Db 243 ThrThrIleTyrProAlaSerGlyGlySerAspTrpAlaTyrAspGlnGlyLys 262
QY 1044 TATTCGTTTATATC-----1057
Db 263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282
QY 1058 -----AAACCCACCTGAGAGCTTTGCGCTCTCTCTAAATAGTTCGCGATGTC 1111
Db 283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIleLysTyrValThrSerTyrVal 302
QY 1112 ATTAGGAATGTT 1123
Db 303 LeuGluHisLeu 306

RESULT 8
CBPC RAT
ID CBPC RAT STANDARD; PRT; 309 AA.
AC P21951;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase A3).
GN CPA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA Neureath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme activity within mast cell granules."
RL Biochemistry 30:648-655(1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A38395; A38395.

DR HSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 68
FT METAL 71
FT METAL 71
FT METAL 196
FT ACT_SITE 248
FT ACT_SITE 270
FT ACT_SITE 270
FT DISULFID 65
FT DISULFID 137
FT VARIANT 1
SQ SEQUENCE 309 AA; 35786 MW; 20330FABC3EE83EF CRC64;
Alignment Scores:
Pred. No.: 1.46e-42 Length: 309
Score: 567.50 Matches: 120
Percent Similarity: 54.79% Conservative: 46
Best Local Similarity: 39.60% Mismatches: 80
Query Match: 20.44% Indels: 58
DB: 1 Gaps: 6
US-09-980-881A-1 (1-1573) x CBPC_RAT (1-309)
QY 375 TATGAACAGTATCACTCACTAAATGAATCATTTCTTTGATAGAAATTTATACTGAGAG 434
Db 8 TyrAlaLysTyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLys 27
QY 435 CATCTCGATATGCTTACAAATCCACATTTGATGATCTCTCATTTGAGAGTACCCACTTAT 494
Db 28 HisProGluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyr 47
QY 495 GTTTTAAAGTTTCTCGAAAGAACAAACAGCCAAAATGCCATATGATGATGATGATGATG 554
Db 48 ValLeuLysIle---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66
QY 555 ATCCATGCCAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 602
Db 67 IleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86
QY 603 -----GGCCATAT-----611
Db 87 LysSerTyrGlyLysAsnAsnIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106
QY 612 -----CGAATGCGAGA 623
Db 107 LeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArg 126
QY 624 AAGAACCGTTTCTTCTATGCGAACAAATCATTTGCGAAGACAGACCTGGAATAGCAACTTT 683
Db 127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146
QY 684 GTCTCCAAACACTGGTGTAGGAAGAGTGCATCCAGTCTCTCTCTCTCTCTCTCTCTCTCT 743
Db 147 ---AspValSerTrpAspSerSerProAsnThrAspAsnProCysLysLeuSerValTyrArg 165
QY 744 GGACTTTATCTGAGTCAGACACAGAGTGAAGCAGTGGCTAGTCTTCTGAGAAGAAAT 803
Db 166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185
QY 804 ATCAACACAGATTAAGCATATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTC 863
Db 186 LeuAsnSerIleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205
QY 864 TATTCCTATACAGAAAGTAAAGCAAGACCTGAGAACTGCTCTAGTAGGAGAGTAA 923
Db 206 TyrGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLysValAlaArgIle 225

Db 75 VallysAlaPheLeuGluTyrHisGlyIleSerTyrGluIleMetIleGluAspValGln 94
 Qy 315 GATCTTATT-----CAACAGCAGATTTCACAGACAGATGTCAGCCCGAGCCTCC 365
 Db 95 LeuLeuLeuAspGluGluTyrHisGlnMetSerAlaPheGlnAlaArgAlaLeuSerThr 114
 Qy 366 GCATCGTAC--TATGAACAGTATCACTCACTAAATGAAATCTATCTTGGATAGATT 422
 Db 115 AspSerPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrGluPheMetAspLeu 134
 Qy 423 ATAAGTGGAGGATCTGATATGCTTACAAAAATCCACATTCGATCGCTCTATTTGAGAAG 482
 Db 135 LeuValAlaGluHisProGlnLeuValSerIleGlnIleGlyAsnThrPheGluGly 154
 Qy 483 TACCACCTCTATGTTTAAAGTTTCTGGAAGAACAAACAGCCAAATATGTCATATGG 542
 Db 155 ArgProIleHisValLeuLysPheSer---ThrGlyGlyThrAsnArgProAlaIleTyr 173
 Qy 543 ATTGACTGTGGAATCCATCCAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 602
 Db 174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193
 Qy 603 GGCCAT----- 608
 Db 194 LysIleThrLysAspTyrGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213
 Qy 609 -----AAT 611
 Db 214 AspilePheLeuGluIleValThrAsnProAspGlyPheAlaTyrThrHisLysThrAsn 233
 Qy 612 CGAATGTGGAGAAGAACCGTTCTTCTATCGAACCAATCATTCATCGAACAGACCTG 671
 Db 234 ArgMetTrpArgLysThrArgSerHisThrGlnGlySerLeuGlyValGlyValAspPro 253
 Qy 672 AATAGCAACTTGTCTCCAAACACTGTGTGAGGAGGTGCATCCAGTCTCTCATGCTCG 731
 Db 254 AsnArgAsnTrp---AspAlaGlyPheGlyWeAlaGlyAlaSerSerAsnProCysSer 272
 Qy 732 GAAACCTACTGTGAGCTTTATCTCTGAGTCAGAACAGAGTGAAGCAGTGGCTAGTTTC 791
 Db 273 GluThrTyrArgGlyLysPheProAsnSerGluValGluValLysSerIleValAspPhe 292
 Qy 792 TTGAGAAGAAATCAACAGATTAAGCATACATCAGCATGATCATCATCATCATCATCAT 851
 Db 293 ValThrSerHisGlyAsn---IleLysAlaPheIleSerIleHisSerTyrSerGlnLeu 311
 Qy 852 ATAGTGTTCATATCTCTATACAGAGTAAAGCAAGCAACAGCATGAGGAGTCTCTA 911
 Db 312 LeuLeuTyrProTyrGlyTyrThrSerGluProAlaProAspGlnAlaGluLeuAspGln 331
 Qy 912 GTAGCCAGTGAAGCAGTTCGTCTATTGACAAACTAGTAAATAACAGGATATACAT 971
 Db 332 LeuAlaLysSerAlaValThrAlaLeu---ThrSerLeuHisGlyThrLysPheLysTyr 350
 Qy 972 GGCCATGGCTCAGAACTTATACCTAGTCTGAGGTGGGAGGAGGAGGAGGAGGAGGAG 1031
 Db 351 GlySerIleLeuAspThrIleTyrGlnAlaSerGlySerThrIleAspTrpThrTyrSer 370
 Qy 1032 TTGGGCAATCAATTCGTTTACATCAAC----- 1061
 Db 371 GlnGlyIleLysTyrSerPheThrPheGluLeuArgAspThrGlyLeuArgGlyPheLeu 390
 Qy 1062 CCACCTGTAGAGAGCTTTTGGCGCTCTCTAAATAGCTTGGCATGTATTAGGAATG 1121
 Db 391 LeuProAlaSerGlnIleIleProThrAlaGlu-----ThrTrpLeuAlaLeu 407
 Qy 1122 TTTAAGCCCTGATTTTATCATCTCTCCGAT 1157
 Db 408 LeuThrIleMetAspHisThrValLysHisProTyr 419
 RESULT 12
 ID CPB2_HUMAN STANDARD; PRT; 417 AA.

AC P48052; Q96A12; Q96Q03;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
 GN CPA2.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
 RC TISSUE=Pancreas;
 RX MEDLINE=95204457; PubMed=7896805;
 RA Catusas L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
 RT Billster M.;
 RT "The sequence and conformation of human pancreatic
 RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and
 RT three-dimensional model.";
 RL J. Biol. Chem. 270:6651-6657(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=98111000; PubMed=9450539;
 RA Reverter D., Garcia-Saez I., Catusas L., Vendrell J., Coll M.,
 RA Aviles F.X.;
 RT "Characterisation and preliminary X-ray diffraction analysis of human
 RT pancreatic procarboxypeptidase A2.";
 RL FEBS Lett. 420:7-10(1997).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=98046021; PubMed=9384570;
 RA Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
 RT "The three-dimensional structure of human procarboxypeptidase A2.
 RT Deciphering the basis of the inhibition, activation and intrinsic
 RT activity of the zymogen.";
 RL EMBO J. 16:6908-6913(1997).
 CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
 CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
 CC residues.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC

DR EMBL; U19977; AAA74425.1; --
 DR EMBL; BC007009; AAH07009.1; --
 DR EMBL; BC014571; AAH14571.1; --
 DR EMBL; BC015140; AAH15140.1; --
 DR PIR; A56171; A56171.
 DR PDB; 1AYE; 13-JAN-99.
 DR PDB; 1D7D; 12-JUL-00.
 DR PDB; 1O6X; 30-JAN-03.
 DR MEROPS; M14.002; --
 DR Genew; HGNC:2297; CPA2.
 DR MIN; 60688; --
 DR GO; GO:0004182; F:carboxypeptidase A activity; TAS.
 DR GO; GO:0007039; P:vacuolar protein catabolism; TAS.
 DR InterPro; IPR00146; Propep_M14.
 DR InterPro; IPR00834; Zn_carbopept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carbopept; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR SMART; SM00631; Zn_pept; 1.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 DR Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
 KW 3D-structure.
 FT SIGNAL 1 16
 FT PROPEP 17 112
 FT CHAIN 113 417
 FT METAL 177 177
 FT METAL 180 180
 FT METAL 304 304
 FT ACT_SITE 378 378
 FT DISULFID 246 269
 FT DISULFID 318 352
 FT CONFLICT 80 80
 FT CONFLICT 302 302
 FT CONFLICT 417 AA; 46928 MW; 3D9C073A98700A83 CRC64;
 SQ SEQUENCE 417 AA; 46928 MW; 3D9C073A98700A83 CRC64;
 Alignment Scores:
 Pred. No.: 1.64e-37 Length: 417
 Score: 511.00 Matches: 128
 Percent Similarity: 48.73% Conservative: 64
 Best Local Similarity: 32.49% Mismatches: 136
 Query Match: 18.40% Indels: 66
 DB: 1 Gaps: 11
 US-09-980-881a-1 (1-1573) x CPB2_HUMAN (1-417)
 QY 72 CATGCTCTC-----GGTCCAGAGTGGCCAAAGTCTTAGCTCTCTCTAGAAC 122
 Db 13 HisileTyrcysLeuGluThrPheValGlyAspGlnValLeuGluLeuValProSerAsn 32
 QY 123 TCTAGGCAAGTTCAGTCTTACAGAACTTACTACACATATGAGATTGTCTC----- 176
 Db 33 GluGluGlnIleTyrsAsnLeuLeuGluGlnLeuGluAlaGlnGluHisLeuGlnLeuAspPhe 52
 QY 177 TGGCAGCGGTAAACAGCTGACCTTATTGTGAGAAAAAACAAGTCCATTTTGTGTAAT 236
 Db 53 TrpLysSerProThr-----ThrProGluThrAlaHisValArgValpro 68
 QY 237 GCATCTGATGCGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTGC 296
 Db 69 PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaTyrSerIle 88
 QY 297 TTGCTGGCAGCGTGGAGATCTTATTCACAGCAGATTTCCACGACACAGTCAGCCCC 356
 Db 89 MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg 108
 QY 357 CGAGCCTCCGATCG-----TACTATGAACAGATATCACTCACTAAATGAATCTAT 407
 Db 109 ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaTyrHisThrLeuGluGluIleSer 128
 QY 408 TCTTGGATAGAAATTTAACTAGGAGCGCATCTTGATATGCTTACAAAATCCATTTGGA 467
 Db 129 GlnGluMetAspAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly 148

QY 468 TCCTCATTTTGAGAGTACCACCTCTATGTTTAAAGTTTCT---GGAAAAGAACAAACA 524
 Db 149 SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyGlyAsp----- 166
 QY 525 GCCAAAATGCCCATATGGATTGACTGTGGAAATCCATGCCAGAGATGGATCTCTCTGCT 584
 Db 167 ---LysProAlaIleIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla 185
 QY 585 TTCTGCTTGTGGTTTCATAGGC----- 605
 Db 186 ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205
 QY 605 ----- 605
 Db 206 SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal 225
 QY 606 -----CATATCGAATGTGGAGAAAGACCGTTCTTTCTATGCGAACAAATCAT 653
 Db 226 PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLysValSerGlySerLeu 245
 QY 654 TGCATCGGAACAGACCTGATAGCACTTTGTCTCCAAACACACTGGTGTGAGGAAGTGCA 713
 Db 246 CysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla 264
 QY 714 TCAGATTCTCTCATGCTCGGAACCTACTGTGGACTTTATCTCTGAGTCAGAACCAAGTG 773
 Db 265 SerSerAsnProCysSerAspSerTyrHisGlyProSerAlaAsnSerGluValGluVal 284
 QY 774 AAGCAGATGGCTAGTTTCTTTGAGAAGAAATATCAACCAAGATTAAAGCATACATCAGCATG 833
 Db 285 LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleThrLeu 303
 QY 834 CATTCATCTCCAGCATATAGTGTTCATATCTCTATACAGAAATTAACGAAGCAAGAC 893
 Db 304 HisSerTy-SerGlnLeuLeuMetPheProTyrcysLeuLeuMetPheProTyrcysLeuLeuAspAsp 323
 QY 894 CATGAGGAAGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAATAGTAGTAA 953
 Db 324 PheAspGluLeuSerGluValAlaGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis 342
 QY 954 AATACCAAGGTATACACATGCGCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGG 1013
 Db 343 GlyThrLysTyrsValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySer 362
 QY 1014 GAGATTGATCTATGATTGGGCATCAATATTCGTTTACATCAAC----- 1061
 Db 363 IleAspTrpSerTyrAspTyrGlyIleLysTyrSerPheAlaPheGluLeuArgAspThr 382
 QY 1062 -----CCACCTGTAGAGAAGCTTTTGCCG 1085
 Db 383 GlyArgTyrGlyPheLeuLeuProAlaArgGlnIleLeuPro 396
 RESULT 13
 CBP1_HUMAN
 ID CBP1_HUMAN STANDARD; PRT; 419 AA.
 AC P15085; Q9BS67;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPA1 OR CPA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=93038569; PubMed=1417781;
 RA Cateasus L., Villegas V., Pascual R., Aviles F.X.,
 RA Wicker-Planquart C., Puigserver A.;
 RT "cDNA cloning and sequence analysis of human pancreatic


```
QY 597 -----TTCCATA----- 602
Db 212 LeuGluLysMetAspIlePheLeuLeuProValAlaAsnProAspGlyTyrValTyrThr 231
QY 603 ---GGCCATATCAATGTCGAGAAAGAACCGTTCTTCTTCTATCGAACAATCATTCGATC 659
Db 232 GlnThrGlnAsnArgLeuTpaGlyThrArgSerArgAsnProGlySerCysIle 251
QY 660 GGAACAGACCTGAATAGCACTTTGTCTCCAAACACTGTGTGTGAGGAAGTGCATCCAGT 719
Db 252 GlyAlaAspProAsnArgAsnTpr---AsnAlaSerPheAlaGlyLysGlyAlaSerAsp 270
QY 720 TCCTCATGCTCGAAACCTACTGTGACATTATCTCAGTCAGACAGAACAGAGTGAAGGCA 779
Db 271 AsnProCysSerGluValTyrHisGlyProHisAlaAsnSerGluValGluValLysSer 290
QY 780 GTGGCTAGTTTCTTGAGAGCAATATCAACACGATTAAAGCATACATCACATGCAATTC 839
Db 291 ValValAspPheIleGlnLysHisGlyAsn---PheLysGlyPheIleAspLeuHisSer 309
QY 840 TACTCCAGCATATAGTGTGTTTCCATATTCCTATACACAGAACTAAAGCAAGAACCATGAG 899
Db 310 TyrSerGlnLeuLeuMetTyrProTyrGlyTyrSerValLysLysAlaProAspAlaGlu 329
QY 900 GAATGTCTCTAGTAGCAGTGAAGCAGTGTGCTGCTATTCGACAAACTAGTAAATAATACC 959
Db 330 GluLeuAspLysValAlaArgLeuAlaAlaLysAlaLeuAlaSerValSer---GlyThr 348
QY 960 AGGTATACATGCGCCATGCTCAGAAACCTTATCTAGTCTCTCGAGTGGGGAGCAT 1019
Db 349 GluTyrGlnValGlyProThrCysThrValTyrProAlaSerGlySerSerIleAsp 368
QY 1020 TGGATCTATGATTGGGCATCAATATTCGTTTACA 1055
Db 369 TipAlaTyrAspAsnGlyIleLysPheAlaPheThr 380

RESULT 15
CBPZ_SIMVI
ID -CBPZ_SIMVI STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidae;
OC Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gut;
RX MEDLINE=9403864; PubMed=8269093;
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins.";
RL Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: Involved in the digestion of the blood meal.
CC -1- CELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Gut-specific.
CC -1- INDUCTION: By blood meal.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L08481; AAA18531.1; -.
```

```
DR HSP, P48052; 1AYE.
DR MEROPS; M14. UPA; -.
DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF00246; Zn_carboOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 58 58 ZINC (BY SIMILARITY).
FT METAL 61 61 ZINC (BY SIMILARITY).
FT METAL 184 184 ZINC (BY SIMILARITY).
FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 125 148 BY SIMILARITY.
SQ SEQUENCE 304 AA; 34849 MW; 2E6E3FF8A6AA9144 CRC64;

Alignment Scores:
Pred. No.: 1.84e-28 Length: 304
Score: 409.00 Matches: 98
Percent Similarity: 49.68% Conservative: 55
Best Local Similarity: 31.82% Mismatches: 95
Query Match: 14.73% Indels: 60
DB: 1 Gaps: 10

US-09-980-881A-1 (1-1573) x CBPZ_SIMVI (1-304)
QY 381 CAGTATCACTCACTAAATGAAATCTATTCTTGATAGAAATTTATACTGAGAGGCATCCT 440
Db 1 GlnTyrHisThrLeuProGluIleTyrSerTrpLeuAspArgLeuValGlnGluHisPro 20
QY 441 GATATGCTTCAAAAATCCACATTCGATCCTCATTTGAGAAGTACCCTCTATGTTTAA 500
Db 21 GluHisValGluProValValGlyLysSerTyrGluGlyArgGluIleAeGlyVal 40
QY 501 AAGTITTCGGAAGAAACAAACAGCCAAAATGCCATATGATTCATGTCGGAATCCAT 560
Db 41 LysValSerTyrLysLys-----GlyAsnProValValMetValGluSerAsnIleHis 58
QY 561 GCCAGAGAATGATCTCTCTGCT-----TTCCTGCTTG----- 593
Db 59 AlaArgGluTrpIleThrAlaAlaThrThrTyrLeuLeuAsnGluLeuLeuThrSer 78
QY 594 -----TGGTTTCATA----- 602
Db 79 LysAsnSerThrIleArgGluMetAlaGluAsnTyrAspTrpTyrIlePheProValThr 98
QY 603 -----GGCAT-----AATCGAATGTGGAGAAGAACCCCTTCT 635
Db 99 AsnProAspGlyTyrValTyrThrHisThrThrAspArgMetTrpArgLysThrArgSer 118
QY 636 TTCATGCGAACAAATCATTCGACACAGACCTGAAATAGCAACTTTGTCTCCAAACAC 695
Db 119 ProAsnProAspSerLeuCysAlaGlyThrAspProAsnArgAsnTrp---AsnPheHis 137
QY 696 TGGTGTGAGGAAGTGCATCCAGTTCCTCATGTCGAAACCTACTGTGAGACTTTATCCT 755
Db 138 TrpMetGluGlnGlyThrSerSerArgProCysThrGluThrTyrGlyGlyLysLysAla 157
QY 756 GAGTCAGAACACAGAGTGAAGGAGTGGCTAGTTCTTCGAGAAGAAATATCAACACAGATT 815
Db 158 PheSerGluValGluThrArgSerPheSerPheLeuLysThrLeuLysGlyGlnIle 177
QY 816 AAAGCATATACATCAGCATGCTATCATCTCCAGCATATAGTGTTCCTATTCCTATACA 875
Db 178 LysValTyrLeuAlaPheHisSerTyrSerGlnLeuLeuLeuPheProTyrGlyHisThr 197
QY 876 CGAAGTAAAGCAAGAACCATGAGCACTGCTCTAGTAGCCAGTGAACAGTTCGTGCT 935
Db 198 CysGlnHisThrTyrAsnHisAspLeuGlnAlaIleGlyAspAlaAlaArgSer 217
QY 936 ATGCAAAATAGTAAATAATACAGGTATACACATGCGCATGCGCTCAGAACCTTTATAC 995
Db 936 ATGCAAAATAGTAAATAATACAGGTATACACATGCGCATGCGCTCAGAACCTTTATAC 995
```

```
Db      218 LeuAlaGln---ArgTyrGlyThrAspTyrThrValGlyAsnIleTyrAspAlaIleTyr 236
Qy      996 CTAGCTCTGAGAGTGGGACGATTGGATCTAT-----GATTGGGCATCAAAATAT 1046
Db      237 ProAlaSerGlyGlySerMetAspIleAlaTyrAspThrLeuAspIleProIleAlaTyr 256
Qy      1047 TCGTTTACATCAAAACCCA-----CCTGTAGAGAAGCTT 1079
Db      257 ThrTyrGluLeuArgProArgAspGlyTrpAsnGlyPheGlnLeuProAlaAsnGlnIle 276
Qy      1080 TTGCCG-----CTGTCTCTAAATAGCTTGGCATGTCA 1112
Db      277 IleProThrGlyGluGluThrValAspSerValThrIleLeuLysGluSerArgArg 296
Qy      1113 TTAGGAATGTTTAATGCCCTGAT 1136
Db      297 LeuGlyTyrPheAsnThrSerAsp 304
```

Search completed: January 1, 2004, 18:53:49
Job time : 40.6016 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:10:36 ; Search time 78.8575 Seconds
(without alignments)
10294.941 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 aaaaattctgttgatg.....aaaaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2/1/USPTO.spool/US09980881/runat_31122003_135748_24954/app_query.fasta_1.2958
-DB=SPTRMBL_23 -OFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pept -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN 1 1 190 @runat_31122003_135748_24954 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvrius:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1885	67.9	423	4	Q15114	Q15114 homo sapien

2	1884	67.8	360	4	Q9P2Y6	Q9P2Y6 homo sapien
3	1880	67.7	423	4	Q961Y4	Q961Y4 homo sapien
4	1592.5	57.3	422	11	Q9JHH6	Q9JHH6 mus musculus
5	1586.5	57.1	422	11	Q9QZF0	Q9QZF0 mus musculus
6	1568.5	56.5	422	11	Q9EQV9	Q9EQV9 rattus norv
7	1029	37.1	198	4	Q9NTI8	Q9NTI8 homo sapien
8	714	25.7	416	13	Q9PUF2	Q9PUF2 bothrops ja
9	703.5	25.3	408	13	Q8AXN3	Q8AXN3 paralichthy
10	646	23.3	416	6	Q9XSP3	Q9XSP3 sus scrofa
11	642.5	23.1	412	11	P97597	P97597 rattus norv
12	633.5	22.8	437	4	Q8NA4T0	Q8NA4T0 homo sapien
13	625	22.5	434	13	Q8AVS2	Q8AVS2 xenopus lae
14	581	20.9	419	13	Q8UUK1	Q8UUK1 gallus gall
15	570	20.5	374	4	Q8IVL8	Q8IVL8 homo sapien
16	556	20.0	436	11	Q8R4H4	Q8R4H4 mus musculus
17	534	19.2	419	13	Q8AXN4	Q8AXN4 paralichthy
18	520	18.7	436	4	Q8WXQ8	Q8WXQ8 homo sapien
19	516.5	18.6	419	6	Q9TV85	Q9TV85 sus scrofa
20	507.5	18.3	420	11	Q8BMK6	Q8BMK6 mus musculus
21	504	18.1	262	11	Q8BVD0	Q8BVD0 mus musculus
22	490	17.6	452	13	Q8UW45	Q8UW45 fugu rubrip
23	484	17.4	247	4	Q96KZ9	Q96KZ9 homo sapien
24	461	16.6	428	5	Q8IGC4	Q8IGC4 drosophila
25	460.5	16.6	419	13	Q8AXN5	Q8AXN5 paralichthy
26	460	16.6	424	5	Q9VL87	Q9VL87 drosophila
27	459	16.5	1132	5	Q9W475	Q9W475 drosophila
28	451.5	16.3	422	5	Q9VLZ2	Q9VLZ2 drosophila
29	440.5	15.9	226	4	Q8TDE8	Q8TDE8 homo sapien
30	436	15.7	584	5	Q9TZH1	Q9TZH1 caenorhabdi
31	434.5	15.6	351	5	Q8TQ1	Q8TQ1 drosophila
32	434	15.6	351	4	Q8NA08	Q8NA08 homo sapien
33	418	15.1	445	5	Q9W478	Q9W478 drosophila
34	416.5	15.0	488	5	Q9BL88	Q9BL88 caenorhabdi
35	414.5	14.9	416	5	Q9VRZ3	Q9VRZ3 drosophila
36	407.5	14.7	419	5	Q9VLZ1	Q9VLZ1 drosophila
37	403	14.5	430	5	Q8SZP6	Q8SZP6 drosophila
38	402	14.5	424	5	Q9VS66	Q9VS66 drosophila
39	402	14.5	430	5	Q9VL86	Q9VL86 drosophila
40	400.5	14.4	540	5	Q9TZH2	Q9TZH2 caenorhabdi
41	399.5	14.4	427	5	Q9U9K2	Q9U9K2 aedes aegypt
42	397	14.3	323	5	Q19121	Q19121 caenorhabdi
43	396	14.3	453	5	Q9V342	Q9V342 drosophila
44	386.5	13.9	423	5	Q61532	Q61532 drosophila
45	382.5	13.8	581	5	Q23318	Q23318 caenorhabdi

ALIGNMENTS

RESULT 1
Q15114
ID Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PCPB protein.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; M75105; AAA60042.1; -;
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.009; -;
DR Genew; HGNC:2300; CPB2.

Percent Similarity: 98.89%		Conservative: 2
Best Local Similarity: 98.33%		Mismatches: 4
Query Match: 67.84%		Indels: 0
DB: 4		Gaps: 0
US-09-980-881A-1 (1-1573) x Q96IY4 (1-360)		
QY	18	ATGAAGCTTTGCAGCTTGCAGTCTGTACCCATTGTTCTCTCTGTGAGCAGCATGTC 77
Db	1	MetLysLeuCySerLeuAlaValLeuValProileValLeuPheCysGluGlnHisVal 20
QY	78	TTGCGGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTCTAGAACCTCTAGCGAAGTTCAA 137
Db	21	PheAlaPheGlnThrGlyGlnValLeuAlaLeuProargThrSerArgGlnValGln 40
QY	138	GTCTTACAGAACTTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 137
Db	41	ValLeuGlnAenLeuThrThrThrTyrrGluileValLeuTrpGlnProValThrAlaasp 60
QY	198	CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGAAATGTG 257
Db	61	LeuileValLysLysGlnValHisPheValAenAlaSerAspValAspAenVal 80
QY	258	AAAGCCCATTTAAATGTAGCGGAATTCATGCAAGTCTTGTGGCAGCGGTAAACAGCTGAC 317
Db	81	LysAlaHisLeuAenValSerGlyIleProCysSerValLeuLeuAlaAspValGluasp 100
QY	318	CTTATTCAACAGCAGATTTCCACACGACAGTCCAGCCCGCCGCTCCGCATCGTACTAT 377
Db	101	LeuileGlnGlnGlnLeuSerAenAspThrValSerProArgAlaSerAlaSerTyrr 120
QY	378	GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATACTGAGAGCAT 437
Db	121	GluGlnTyrrHisSerLeuAenGluileTyrrSerTrpIleGluPheileThrGluArgHis 140
QY	438	CTGTGATATGCTTACAAAAATCCACATTCATTCATTTGAGAAAGTACCCACTCTATGTT 497
US-09-980-881A-1 (1-1573) x Q96IY4 (1-423)		
QY	18	ATGAAGCTTTGCAGCTTGCAGTCTGTACCCATTGTTCTCTCTGTGAGCAGCATGTC 77
Db	1	MetLysLeuCySerLeuAlaValLeuValProileValLeuPheCysGluGlnHisVal 20
QY	78	TTGCGGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTCTAGAACCTCTAGCGAAGTTCAA 137
Db	21	PheAlaPheGlnSerGlyGlnValLeuAlaLeuProargThrSerArgGlnValGln 40
QY	138	GTCTTACAGAACTTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGAC 197
Db	41	ValLeuGlnAenLeuThrThrThrTyrrGluileValLeuTrpGlnProValThrAlaasp 60
QY	198	CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGAAATGTG 257
Db	61	LeuileValLysLysGlnValHisPheValAenAlaSerAspValAspAenVal 80
QY	258	AAAGCCCATTTAAATGTAGCGGAATTCATGCAAGTCTTGTGGCAGCGGTAAACAGCTGAC 317
Db	81	LysAlaHisLeuAenValSerGlyIleProCysSerValLeuLeuAlaAspValGluasp 100
QY	318	CTTATTCAACAGCAGATTTCCACACGACAGTCCAGCCCGCCGCTCCGCATCGTACTAT 377
Db	101	LeuileGlnGlnGlnLeuSerAenAspThrValSerProArgAlaSerAlaSerTyrr 120
QY	378	GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATACTGAGAGCAT 437
Db	121	GluGlnTyrrHisSerLeuAenGluileTyrrSerTrpIleGluPheileThrGluArgHis 140
QY	438	CTGTGATATGCTTACAAAAATCCACATTCATTCATTTGAGAAAGTACCCACTCTATGTT 497

```
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Qy 498 TTAAGCTTTCTGGAAGACAAACAGCCAAATGCCATATGATGATGCTGGATC 557
Db 161 LeuLysValSerGlyLysGluGlnAlaLysAsnAlaIleTrpIleAspCysGlyLe 180
Qy 558 CATCCAGAGAAATGGATCTCTCCCTGCTTCTGCTGTTGCTGTTCTAGGCGCAT 608
Db 181 HisAlaArgGlnTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
Qy 608 ----- 608
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValMet 220
Qy 609 -----AATCGAATGTGAGAAAG 626
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
Qy 627 AACCGTTCTTCTATGCGAACAATCATTTGCATCGGAACAGACCTGAATAGCACTTTGTC 686
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
Qy 687 TCACAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 746
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
Qy 747 CTTTATCTGAGTCAGAACAGCAAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATC 806
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 300
Qy 807 AACCCAGATTAAGCATACATCAGCATGCTTCTACTCCAGCATATAGTGTTCATAT 866
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
Qy 867 TCCTATACAGAGTAAAGCAAGCAACAGCATGAGCACTGCTCTAGTAGCCAGTGAAGCA 926
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
Qy 927 GTTCGCTGTTATGCAAAACTAGTAAATAACAGGTATACACATGCCCAGTGCAGAA 986
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
Qy 987 ACCTTATACCTAGTCTCTCGAGGTGGGACGATGATGATGATGATGATGATGATGAT 1046
Db 361 ThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGlyIleLysTyr 380
Qy 1047 TCGTT-----TAC 1054
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
Qy 1055 ATCAAAACCCAGCTTAGAGAGCTTTTGGCGCTGCTCTCTCTCTCTCTCTCTCTCTCT 1114
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLysIleAlaTrpHisValIle 420
Qy 1115 AGGAATGTT 1123
Db 421 ArgAsnVal 423
RESULT 4
Q9JHH6 PRELIMINARY; PRT; 422 AA.
AC Q9JHH6
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE (1110032P04Rik protein).
GN CPB2 OR TAFI OR 1110032P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN (1)
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20201996; PubMed=10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor.";
RL Thromb. Haemost. 83:297-303 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kieda K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Scahill F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Havaehizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR EMBL; AB021968; BAB03402.1; -
DR EMBL; AF164524; AAF62385.1; -
DR EMBL; AK004045; BAB23141.1; -
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00831; Zn_pept; 1.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

Alignment Scores:
Pred. No.: 1 63e-151 Length: 422
Score: 1592.50 Matches: 309
Percent Similarity: 79.15% Conservative: 25
Best Local Similarity: 73.22% Mismatches: 33
Query Match: 57.35% Indels: 56
DB: 11 Gaps: 3

US-09-980-881A-1 (1-1573) x Q9JHH6 (1-422)
Qy 18 ATCAAGCTTTGACGCTTGCAGCTTGCAGTCTGTGTACCATGTTCTTCTGTGTGAGCAGCATGTC 77
Db 1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleLeuTyr---GluGlnHisGly 19
Qy 78 TTCGCGTTCCAGATGCCCAAGTCTAGCTGCTCTTCTCTAGACCTCTAGCGAGTTCAA 137
Db 20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39
Qy 138 GTTCTACAGAAATCTTACTCAACATATGAGATTGTTCTCTGGCGCGCGGTAAACAGCTGAC 197
:::|||||
```

```
Db 40 LeuLeuGlnAsnLeuThrThrThrThrTyTyTyGluValValLeuThrGlnProValThrAlaGlu 59
Qy 198 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 257
Db 60 PheileGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79
Qy 258 AAAGCCCATTTAAATGTAGCGGAATTCATGCGAGTCTTGTCTGGCAGACGTGGAAGAT 317
Db 80 LysAlaHisLeuAsnValSerArgileProPheAsnValLeuMetAsnValGluAsp 99
Qy 318 CTTATTCAACAGCAGATTTCCAACGACACATGTCAGCCCGAGCCTCCGATCGTACTAT 377
Db 100 LeuileGluGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyTy 119
Qy 378 GAACAGTATCACTCACTAAATCTATCTTGGATAGATTAATTAACGTAGAGGCAT 437
Db 120 GluGlnTyHisSerLeuAsnGluileTySerTrpileGluValileThrGluGlnHis 139
Qy 438 CCTGATATGCTTCAAAAATCCCATGATGATCCCTCAATTCAGAGATACCCACTCTATGTT 497
Db 140 ProAspMetLeuGlnLysileTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 159
Qy 498 TTAAGGTTTCTGGAAGAAACAACAGCCAAAATGCCATATGCGATTGATGCTGGAATC 557
Db 160 LeuLysValSerGlyLysGluGlnArgLleLysAsnAlaileTrpileAspCysGlyile 179
Qy 558 CATGCCAGAGATGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 608
Db 180 HisAlaArgGluTrpileSerProAlaPheCysLeuTrpPheileGlyTyTyValThrGln 199
Qy 608 ----- 608
Db 200 PheHisGlyLysGluAsnLeuTyThrArgLeuLeuArgHisValAspPheTyTyMet 219
Qy 609 -----AATCGAATGTGAGAAAG 626
Db 220 ProValMetAsnValAspGlyTyTyAspTyTyThrTrpLysLysAsnArgMetTrpArgLys 239
Qy 627 AACCGTTCTTTCTATGCGAAACAATCATTTGTCATCGGAACAGACCTGAATAGCAATTTGTC 686
Db 240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259
Qy 687 TCCAAACACTGGTGTAGGAGGTGCATCCAGTTCCTCTGCTCGGAAACCTACTGTGGA 746
Db 260 SerLysHisTrpCysGluLysGlyAlaSerSerSerCysSerGluThrTyTyCysGly 279
Qy 747 CTTTATCTCTGAGTACAGAACAGAGTGAAGTCAAGTGTCTTCTTGTGAGAAGAAATATC 806
Db 280 LeuTyTrpGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnile 299
Qy 807 AACAGATTAAGCATATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 866
Db 300 AspHisileLysAlaTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 319
Qy 867 TCTTATCAGCAAGTAAAGCAAGACCATGAGAACTGCTCTAGTAGCAGTGAAGCA 926
Db 320 SerTyAsnArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 339
Qy 927 GTTCGTCTATGACAAAACCTAGTAAATACCAAGTATGATGATGATGATGATGATGATGATG 986
Db 340 ValArgAlaileGluSerileAsnLysAsnThrArgTyTyThrHisGlySerGlySerGlu 359
Qy 987 ACCTTATACCTAGTCTCTGAGGTGGGACGATGATGATGATGATGATGATGATGATGATGAT 1046
Db 360 SerLeuTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 379
Qy 1047 TCGTT-----TAC 1054
Db 380 SerPheThrileGluLeuArgAspThrGlyArgTyTyTyTyTyTyTyTyTyTyTyTyTyTy 399
Qy 1055 ATCAAAACCCAGTGTAGAGACTTTTGGCGTCTCTCTAAATAGCTTGGCATGTCATT 1114
Db 400 IleLysProThrCysAlaGluAlaLeuAlaAlaileSerLysileValTrpHisValile 419
```

```
Qy 1115 AGCAAT 1120
Db 420 ArgAsn 421
```

RESULT 5

```
ID Q9QZF0 PRELIMINARY; PRT; 422 AA.
AC Q9QZF0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J;
RA He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF186188; AAF00528.1; -.
DR HSSP; P00730; SCPA.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxeopt.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxeopt; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;
```

Alignment Scores:

```
Pred. No.: 6,57e-151 Length: 422
Score: 1586.50 Matches: 308
Percent Similarity: 78.91% Conservative: 25
Best Local Similarity: 72.99% Mismatches: 34
Query Match: 57.13% Indels: 56
DB: 11 Gaps: 3
```

US-09-980-881A-1 (1-1573) x Q9QZF0 (1-422)

```
Qy 18 ATGAAGCTTTGACAGCTTGACGTCCTTGTACCCATTTGTTCTTCTGTGAGCAGCATGTC 77
Db 1 MetLysLeuHisGlyLeuGlyileLeuValAlaileLeuTyTyTyTyTyTyTyTyTyTyTyTy 19
Qy 78 TTCGCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTTAGGCAAGTTCAA 137
Db 20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39
Qy 138 GTTCTACAGAACTTACTACACATATGAGATTTGTTCTCTGGCAGCGCGTAACAGCTGAC 197
Db 40 LeuLeuGlnAsnLeuThrThrTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 59
Qy 198 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 257
Db 60 PheileGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79
Qy 258 AAAGCCCATTTAAATGTAGCGGAATTCATGCGAGTCTTGTCTGGCAGACGTGGAAGAT 317
Db 80 LysAlaHisLeuAsnValSerArgileProPheAsnValLeuMetAsnValGluAsp 99
Qy 318 CTTATTCAACAGCAGATTTTCCAACGACACATGTCAGCCCGAGCCTCCGATCGTACTAT 377
Db 100 LeuileGluGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyTy 119
Qy 378 GAACAGTATCACTCACTAAATGAAATCTATCTTGGATAGAAATTTTAACTGAGAGCAT 437
Db 400 PheileGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 419
```



```
QY 609 -----AATCGAATGTGGAGAAAG 626
Db 220 ProValMetAsnValAspGlyTyrAspTyrThrTrpIysAsnArgMetTrpArgLys 239
QY 627 AACGGTCTTCTATCGAACAATCATTTGCATCGGAACACCTGGAATACCACTTTGTC 686
Db 240 AsnArgSerValHisMetAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259
QY 687 TCAAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTCGTCCGAAACCTACTCTGGA 746
Db 260 SerLysHisTrpCysGluLysGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 279
QY 747 CTTATCTCTAGTCAGAACAGAGTGAAGCGAGTGGCTAGTTCTTCTGAGAGAAATATC 806
Db 280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299
QY 807 AACAGATTAAAGCATACATCAGCATGCTCATCTACTCCAGCATATAGTGTTCATAT 866
Db 300 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnIleLeuPheProTyr 319
QY 867 TCCTATACAGAAAGTAAAGCAAGACCATCAGGAAGTGTCTCTAGTAGCCAGTGAAGCA 926
Db 320 SerTyrAsnArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 339
QY 927 GTTCGTCTATTGACAAAACCTAGTAAATAACAGGATATACATGCCCATGGCTCAGAA 986
Db 340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359
QY 987 ACCTTATACCTACTCTCGAGGTGGGAGCATTTGGATCTCATTTGGGCATCAATAT 1046
Db 360 SerLeuTyrLeuAlaProGlyGlySerAspTrpIleTyrAspLeuGlyIleLysTyr 379
QY 1047 TCGTTTACATC----- 1057
Db 380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgPhe 399
QY 1058 ---AAACCCACCTGTAGAGAGCTTTTGGCGTCTCTCTAAATAAGCTTGGCATGTCAAT 1114
Db 400 IleLysProThrCysAlaGluAlaLeuAlaAlaValSerLysIleAlaTrpHisValIle 419
QY 1115 AGGAAT 1120
Db 420 ArgAsn 421
RESULT 7
Q9PUP2 PRELIMINARY; PRT; 198 AA.
AC Q9PUP2;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE BA139H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137141; CAB92622.1; -.
DR HSSP; P00730; 1CPX.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
FT NON_TER 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;
```

```
Alignment Scores: 9.09e-95 Length: 198
Pred. No.: 1029.00 Matches: 196
Score: 99.49% Conservative: 0
Percent Similarity: 99.49% Mismatches: 1
Best Local Similarity: 99.49% Indels: 0
Query Match: 37.05% Gaps: 0
DB: 4

US-09-980-881A-1 (1-1573) x Q9NTI8 (1-198)
QY 18 ATGAACCTTTGAGCCCTTCAGTCTTCAGTCCCTGATCCCATTTGTTCTCTCTGAGCAGCATGTC 77
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 78 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGTCTCTCTAGAACCTCTAGGCAAGTTCAA 137
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 138 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAAACAGCTGAC 197
Db 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 198 CTTATTGTGAAGAAAACAAAGTCCATTTTGTAAATGCATCTCATGTCGCAATGTG 257
Db 61 LeuIleValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 80
QY 258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGTGTCGACAGCTGGAAGAT 317
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY 318 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAGCCCTCCGATCTACTAT 377
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
QY 378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 437
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 438 CTTGATATGCTTACAAAATCCACATTCATCTCTCTTGTGAGAGTACCACCTCTATGTT 497
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 498 TTAAGGTTTCTGAAAAGAACAAACAGCCAAAATGCCATATGGATTGACCTGGAATC 557
Db 161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 558 CATGCCAGAGATGATCTCTCTCTGCTTGTGTTGTTTCATAGGCCAT 608
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
```

RESULT 8

```
Q9PUP2 PRELIMINARY; PRT; 416 AA.
ID Q9PUP2;
AC Q9PUP2;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAY-2003 (T-EMBLrel. 23, Last annotation update)
DE Carboxypeptidase homolog.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA Murbach A.F.; Hayashi M.A.F.; Camargo A.C.M.;
RT "Screening of Bothrops jararaca pancreas cDNA library.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190274; AAF01344.1; -.
DR HSSP; P09955; INSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
```

DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF02246; Zn_carboOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;

Alignment Scores:
Pred. No.: 6.3e-63 Length: 416
Score: 714.00 Matches: 151
Percent Similarity: 53.94% Conservative: 68
Best Local Similarity: 37.19% Mismatches: 127
Query Match: 25.71% Indels: 61
DB: 13 Gaps: 6

US-09-980-881A-1 (1-1573) x Q9PUF2 (1-416)

QY 60 TTCTGTGACGACGATGCTTCGCGTTCACAGTGGCCAGTTCCTAGCTGCTCTCTCCAG 119
DB 14 PheAlaGluThrThrValHisArgPheAspGlyGluLysValTyrArgValThrProArg 33
QY 120 ACCTCTAGGCAAGTTCAGTCTTACAGATCTTACTACAATATGATGATGATGTTCTCTGG 179
DB 34 AsnGluAspGluValTyrPheLeuAsnTyrLeuAlaAsnIleValGlnValAspPheTip 53
QY 180 CAGCGGTAAACGACGTACCTTATGTGAGAAAAAACAAGTCCATTTTTTGTAAATGCA 239
DB 54 ArgProAspSerValGluLeuValLysAlaGluMetThrValAspPheArgIleGluAla 73
QY 240 TCTGATCTGCACATGGAAGCCCATTTAAATGTGAGCGCAATTCACATGCTGCTTCG 299
DB 74 AspArgCysSerGluValGluSerIleGluGlnSerGlyLeuAsnTyrGluIleLeu 93
QY 300 CTGCGACAGCTGGAGATCTTATTCAACAGAGATTTCCAAACGACACAGTCAGCCCGCA 359
DB 94 IleAspAsnLeuGlnAlaValLeuAspArgGlnLeuAspAsnHisAla-----Arg 110
QY 360 GCCTCCCATCTACTATGACAGTATCACTCACTAATGAATCTATCTTGGATGAGAA 419
DB 111 ThrAlaGlyTyrAsnTyrGluLysTyrAsnSerTrpGluLysIleAspAlaTrpThrAla 130
QY 420 TTTATACTGAGGCGATCTGATATGCTTACAAAAATCCACATTTGATCTCATTTGAG 479
DB 131 AspIleAlaAsnGluAsnProSerLeuValSerArgLeuGlnIleGlyThrThrPheGlu 150
QY 480 AAGTACCCTCTATGTTTAAAGTTTCTGGAAGAACAAACAGCCAAATATGCCATA 539
DB 151 GlyArgProMetProLeuLeuLysVal--GlyLysProGlyValAsnLysLysAlaIle 169
QY 540 TGGATTGACTGTGAATCCATCCAGAGATGATCTCTCTGCTTTCTGCTGTTGTTGTC 599
DB 170 PheIleAspCysGlyPheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPhe 189
QY 600 ATA----- 602
DB 190 ValArgGluAlaValArgThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnLys 209
QY 603 -----GGCCAT 608
DB 210 LeuAspPheTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSerTrpLysGln 229
QY 609 AATCGAATGGGAAAGAACCGTTCTTCTATCGCAACAACTATTCATCCGCAACAGAC 668
DB 230 SerArgMetTrpArgLysThrArgSerValAsnAlaGlySerThrCysIleGlyThrAsp 249
QY 669 CTGAATAGCACTTGTCTCAACACACTGTTGTCAGGAGAGTGCATCCAGTCTCTCATGC 728
DB 250 ProAsnArgAsnPhe----AspAlaAlaTrpCysSerValGlyAlaSerArgAsnProCys 268
QY 729 TCGAAACCTACTGTGAGCTTTATCTGAGTCAGAACCCAGAAAGTGAAGCGAGTGGCTAGT 788

Db 269 SerGluThrTyrCysGlySerLysProGluSerGluLysGluThrLysAlaLeuAlaAsp 288
QY 789 TTCTTCAGAGAATAATCAACAGATTAAAGCATACATCAGCATGATTCATCATCTCCAG 848
Db 289 PheIleArgArgAsnArgSerIleIleGlnAlaTyrLeuThrIleHisSerTyrSerGln 308
QY 849 CATATAGTGTTCCTATATTCCTATATACAGAAAGTAAAGCAACACCATGAGGAACCTGCT 908
Db 309 MetLeuLeuTyrProTyrSerTyrThrTyrAspLeuThrSerAsnAsnLysLysLeuAsn 328
QY 909 CTAGTAGCAGTGAACGAGTTCGTCTATGACAAAACCTAGTAAAATACCAGGTATACA 968
Db 329 SerIleAlaLysGluAlaIleArgGluLeu---LysValLeuPheGlyThrGluTyrThr 347
QY 969 CATGGCCATGGCTCAGAAAACCTTATACCTAGCTCCTCGAGAGTGGGACCATGCGATCTAT 1028
Db 348 TyrGlyProGlyAlaAlaThrIleTyrProAlaAlaGlyLysSerAspAspTrpAlaTyr 367
QY 1029 GATTTCGGCATCAATATTCGTTTACATC----- 1057
Db 368 AspGlnGlyIleLysTyrAlaPheThrPheGluLeuArgAspLysGlyArgTyrGlyPhe 387
QY 1058 -----AAACCCACTGTAGAGAAGCTTTTCCGCTGTCTCTATAA 1096
Db 388 AlaLeuProGluSerGlnIleLysProThrCysGluGluThrMetIleAlaValLysTyr 407
QY 1097 ATAGCTTGGCATGTTCATT 1114
Db 408 IleAlaGluTyrMetLeu 413

RESULT 9
Q8AXN3 PRELIMINARY; PRT; 408 AA.
ID Q8AXN3
AC Q8AXN3; 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Carboxypeptidase B (Fragment).
GN CPB.
OS Paralichthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralichthyidae; Paralichthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RA Srivastava A.S., Kurokawa T., Suzuki T.;
RT "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT A2 and B from the Japanese flounder, Paralichthys olivaceus.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB099302; BAC53789.1; -.
FT NON TER 1
SQ SEQUENCE 408 AA; 45904 MW; 809CAACD9D2DA34B CRC64;

Alignment Scores:
Pred. No.: 7.18e-62 Length: 408
Score: 703.50 Matches: 153
Percent Similarity: 54.11% Conservative: 64
Best Local Similarity: 38.15% Mismatches: 125
Query Match: 25.33% Indels: 60
DB: 13 Gaps: 6

US-09-980-881A-1 (1-1573) x Q8AXN3 (1-408)

QY 84 TTCCAGAGTGGCCAGTTCCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAGTTCTA 143
Db 12 PheGluGlyGluLysValPheArgLeuLysProValPheAspGluHisValThrLeuIle 31
QY 144 CAGAACTTACTACACATATGATGATGTTCTCTGCGAGCCGCGTAACAGCTGACCTTAT 203
Db 32 ArgAspLeuAlaAsnSerIleGluValAspPheTrpArgProGluSerGlnLeuVal 51

```
QY 204 GTGAAGAAAACAAAGTCCATTTTGTGTAATCCATCTCATGTCGACAAATGTGAAGCC 263
Db 52 ThrIleAspIleAspValAlaIleArgValProAlaIleTyrLeuAspMetValTyrThr 71
QY 264 CATTAAATGTGACGGAATTCATGCGAGTCTTGTGCGACAGCGTGAAGATCTATT 323
Db 72 LeuLeuGlnSerAspMetGluHisGluValLeuLeuGluAspLeuGlnSerAlaVal 91
QY 324 CAACAGCAGATTTCCAAACGACAGCAGTCAGCGCCGAGCCTCCGCATCGTACTATGAAC 383
Db 92 AspAlaGluAlaAsp-----LeuLysProSerProArgAlaHisSerTyrThrLys 108
QY 384 TATCACTCACTAAATGAATCTATTCTGTGATAGATTTTATACTGAGAGGCATCTGAT 443
Db 109 TyrAsnAsnLeuAspLysValGlnSerTrpIleAlaSerIleSerSerSerProAsp 128
QY 444 ATGCTTACAAAATCCACATGATGATCTCTCATTTGAGAAGTACCACCTCTATGTTTAAAG 503
Db 129 LeuIleSerLysGlnValIleGlyAsnThrTyrGluGlyArgProMetThrValLeuLys 148
QY 504 GTTCTCGAAAGAACAAACAGCAAGCAAAATGCCATATGCAATGCAATGGAATCCATGCC 563
Db 149 LeuGlyLysLysSerSerSerThrLysProAlaIlePheMetAspCysGlyIleHisAla 168
QY 564 AGAGATGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 602
Db 169 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValLysGluAlaLeuSerThrTyr 188
QY 602 ----- 602
Db 189 GlySerAspSerGlnMetThrSerLeuLeuAsnGlnMetAspValPheValLeuProVal 208
QY 603 -----GGCCAT-----AATCGAATGTGAGAAAGAACCGT 632
Db 209 PheAsnIleAspGlyTyrAspPheThrHisLysSerAsnArgMetTrpArgLysThrArg 228
QY 633 TCTTTCTATCGAACAATCATTCATCGATCGAAGACAGCTGAATAGCAACTTTGTCTCCAA 692
Db 229 SerArgLysSerGlySerSerCysIleGlyAlaAspProAsnArgAsnPhe---AspAla 247
QY 693 CACTGGTGTGAGGAAGTCATCCAGTTCTCTCATGTCGGAACCTACTGTGACTTTAT 752
Db 248 GlyTrpCysSerMetGlyAlaSerAspAsnProCysSerAspThrPheCysGlyTyrThr 267
QY 753 CCTGAGTCAGAACCAAGTGAAGGCGAGTGTGCTTCTTGAGAGAAATATCAACCAG 812
Db 268 ProGluSerGluIleGluValLysAsnValAlaAspPheIleArgAsnArgSerIle 287
QY 813 ATTAAGCATATCAGCATGCAATTCATATCCAGCATATAGTGTTCATATTCCTAT 872
Db 288 IleLysAlaTyrLeuThrValHisSerTyrSerGlnLeuLeuPheProTyrSerTyr 307
QY 873 ACAGAGTAAAGCAAGACCATGAGGAATGCTCTAGTACGACGAGTTCGT 932
Db 308 ThrTyrGlnLeuAlaAlaAspHisSerGluLeuMetLysValAlaGluGlyAlaSerAla 327
QY 933 GCTATTGACAAAATAGTAAAAATACCAAGGTATACATCGCCATGGCTCAGAAACCTTA 992
Db 328 AlaLeu---ArgSerLeuTyrGlyThrTyrThrTyrSerGlyProGlyAlaThrIle 346
QY 993 TACTAGTCTCTGAGTGGGAGCATGATGATCTATGATTTGGGCATCAAAATTCGTT 1052
Db 347 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspLeuGlyValLysTyrSerTyr 366
QY 1053 ACATC----- 1060
Db 367 ThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleLys 386
QY 1061 CCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAATAGCTTGGCATGTCAATAGGAAT 1120
Db 387 ProThrCysGluGluThrMetLeuAlaValLysTyrIleAlaIleValGlnLysAsn 406
QY 1121 GTT 1123
```

```
Db 407 Leu 407
RESULT 10
Q9XSP3 PRELIMINARY; PRT; 416 AA.
ID Q9XSP3
AC Q9XSP3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Procarboxypeptidase B precursor (EC 3.4.17.2).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=99321926; PubMed=10391940;
RA Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA Hersberger C., Aviles F.;
RT "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL J. Biol. Chem. 274:18925-19933 (1999).
DR EMBL; AJ133775; CAB46991.1; -.
DR HSSP; P09955; IPBA.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn carboxypept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn carboxypept; 1.
DR PRINTS; PR00765; CRBOXIPTASEA.
DR SMART; SM00631; Zn pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Hydrolase; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 416 PROCARBOXYPEPTIDASE B.
SQ SEQUENCE 416 AA; 47381 MW; 5FA7FCED7B45A6A CRC64;
Alignment Scores:
Pred. No.: 4,53e-56 Length: 416
Score: 646.00 Matches: 140
Percent Similarity: 53.40% Conservative: 64
Best Local Similarity: 36.65% Mismatches: 134
Query Match: 23.26% Indels: 44
DB: 6 Gaps: 7
US-09-980-881A-1 (1-1573) x Q9XSP3 (1-416)
QY 24 CTTTGACGCTTTCAGTCTTGTACCATTTGTTCTTCTGTGAGCAGCATGTCTTCGCG 83
Db 1 MetLeuAlaPheLeuIleLeuValThrValThrLeuAlaSerAlaHisSerGlyGlu 20
QY 84 ----TTCCAGATGGCCAAAGTTCTAGTGTCTTCTAGAACCTCTAGGCAAGTTCAAGTT 140
Db 21 HisPheGluGlyLysValPheArgValAsnValGluAspGluAsnAspIleSerLeu 40
QY 141 CTACAAATCTTACTACAAATATGATGATTGTTCTTCTGCGAGCGGTAAACAGTACCTT 200
Db 41 LeuHisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGln 60
QY 201 ATTGTGAAGAAAAACAACATCCATTTTGTAAATGCATCTGATCTCGACAAATGTGAAA 260
Db 61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGlu 80
QY 261 GCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTCTGCGACAGCTGGAAGATCTT 320
Db 81 AspPheGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerVal 100
QY 321 ATTCACACAGCATTTCCAAACGACAGCAGTCCGCCCGAGCCTCCGATCTACTATGAA 380
Db 101 LeuGluAlaGlnPheAspSerArgVal-----ArgThrThrGlyHisSerTyrGlu 117
```

```
Qy 381 CAGTATCACTCACTAAATGAATCTATTCTTGATGAGAAATTTATTAAGTGGAGGACCTCT 440
Db 118 LysTyrAsnAsnTrpGluThrIleGluAlaThrThrLysGlnValThrSerGluAsnPro 137
Qy 441 GATATGCTTACAAAATCCACATTCATGATCCCTATTGAGAGTACCCACCTATGTTTAA 500
Db 138 AspLeuIleSerArgThrAlaIleGlyThrThrPheLeuGlyAsnAsnIleTyrLeuLeu 157
Qy 501 AAGTCTTCTGAAAAGAACACACAGCCAAAATGCCATATGATGATGATGATGATGATGAT 560
Db 158 LysVal---GlyLysProGlyProAsnLysProAlaIlePheMetAspCysGlyPheHis 176
Qy 561 GCCAGAGATCGATCTCTCCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
Db 177 AlaArgGluTrpIleSerHisAlaPheCysGlnTrpPheValArgGluAlaValLeuThr 196
Qy 602 -----GQCCAT----- 602
Db 197 TyrGlyTyrGluSerHisMetThrGluPheLeuAsnLysLeuAspPheTyrValLeuPro 216
Qy 603 -----GQCCAT----- 603
Db 217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThr 236
Qy 630 CGTTCTTCTATGCGAACATCATTCATGCGAACAGACCTGATAGTAACTGTTGCTCTCC 689
Db 237 ArgSerThrAsnAlaGlyThrThrCysIleGlyThrAspProAsnArgAsnPhe---Asp 255
Qy 690 AAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTATGCTCGGAAACCTACTGTGGACTT 749
Db 256 AlaGlyTrpCysThrThrGlyAlaSerThrAspProCysAspGluThrTyrCysGlySer 275
Qy 750 TATCCTGAGTCAGAACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 809
Db 276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSer 295
Qy 810 CAGATTAAGCATACATCAGCATCATCTACTCTCCAGCATATGAGTGTTCATATTC 869
Db 296 SerIleLysAlaLysLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSer 315
Qy 870 TATACAGAAAGTAAAGCAAGACCATGAGGAATCTCTCTAGTGGAGTGAAGTGAAGT 929
Db 316 TyrAspTyrLysLeuProGluAsnAsnAlaGluLeuAsnAsnLeuAlaLysAlaVal 335
Qy 930 CGTCTATTGACAAAATAGTAAATACAGGATATACATGCGCATGCGTCAAGAAC 989
Db 336 LysGluLeu---AlaThrLeuTyrGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354
Qy 990 TTATACCTAGCTCTCGGAGTGGGACGATGATGATGATGATGATGATGATGATGATGAT 1049
Db 355 IleTyrProAlaAlaGlyLysSerAspTrpAlaTyrAspGlnGlyIleLysTyrSer 374
Qy 1050 TTTTACA 1055
Db 375 PheThr 376

RESULT 11
P97597 PRELIMINARY; PRT; 412 AA.
AC P97597;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mast cell carboxypeptidase A precursor (fragment).
GN R-CPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A..
RC STRAIN=Sprague Dawley;
RX MEDLINE=97149430; PubMed=8996238;
```

```
RA Lutzelschwab C., Pejler G., Aveskogh M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations.";
RL J. Exp. Med. 185:13-29(1997).
DR EMBL; U67914; AAB48267.1; -.
DR HSP; P09555; INSA.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF02246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Signal.
FT NON TER 1
FT SIGNAL <1 10
FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

Alignment Scores:
Pred. No.: 1,02e-55 Length: 412
Score: 642.50 Matches: 143
Percent Similarity: 51.25% Conservative: 62
Best Local Similarity: 35.75% Mismatches: 136
Query Match: 23.14% Indels: 60
DB: Gaps: 7

US-09-980-881A-1 (1-1573) x P97597 (1-412)
Qy 84 TTCAGAGTGGCCAAAGTCTTAGCTGCTCTTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
Db 16 PheAspArgGluLysValPheArgValLysLeuGlnAspGluLysGlnAlaSerIleLeu 35
Qy 144 CAGAATCTTACTCAACATATGATGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
Db 36 LysAsnLeuThrGlnThrIleGluLeuAspPheTyrTrpProAspAlaIleHisAspIle 55
Qy 204 GTGAAGAAAACAAAGTCCATTTTTTTTGAATGTCATGTCATGTCATGTCATGTCATG 263
Db 56 AlaValAsnMetThrValAspPheArgValThrGluLysGluSerGlnThrIleGlnSer 75
Qy 264 CATTAAATGTGAGCGGAATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 323
Db 76 ThrLeuGluGlnHisLysMetAspTyrGluIleLeuIleAsnAspLeuGlnGluIle 95
Qy 324 CAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCTCCGCTGCTGCTGCTGCTG 383
Db 96 AspLysGlnPhe-----AspValLysGluGluIleAlaGlyArgHisSerTyrAlaLys 113
Qy 384 TATCACTACTAAATGAAATCTATTCTTGATAGATTTATATACTGAGAGGCACTCTGAT 443
Db 114 TyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLysHisProGlu 133
Qy 444 ATGCTTACAAAATCCACATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 503
Db 134 MetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLys 153
Qy 504 GTTCTGGAAGAAAGCAACAAACAGCAAAATGCCATATGCTGCTGCTGCTGCTGCTGCTG 563
Db 154 Ile---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAla 172
Qy 564 AGAGATGATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
Db 173 ArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaLysSerTyr 192
Qy 603 GGCCATAAT----- 611
Db 193 GlyLysAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProVal 212
Qy 612 -----CGAATGTGGAGAAAGACCGT 632
|||||
```

```

Db 213 PheAsnValAspGlyTyrIleTpsSerTpsThrLysAspArgMetTpsArgLysAsnArg 232
Qy 633 TCCTTCTATGGAACAATCAATGTCATCGAAGACAGACCTGAATAGCAACTTTGTCTCCAAA 692
Db 233 SerLysAsnProAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspVal 251
Qy 693 CACTGGTGTGAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGACATTAT 752
Db 252 SerTpsAspSerSerProAsnThrAspAsnProCysLeuSerValTysArgGlyProAla 271
Qy 753 CCTGAGTCAGAACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 812
Db 272 ProGluSerGluLysGluThrLysAlaValThrAsnPhelIleArgSerHisLeuAsnSer 291
Qy 813 ATTAACATACATACATACATACATACATACATACATACATACATACATACATACATACAT 872
Db 292 IleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGlyTyr 311
Qy 873 ACAGAGTAAAGCAAGACCATGAGGAAGTCTCTAGTACGAGTGTGAGGAGGAGGAGGAG 932
Db 312 ThrIleLysLeuProProAsnHisGlnAspLeuLysValAlaArgIleAlaThrAsp 331
Qy 933 GCTATTGACAAACTAGTAAATAATACAGGATATACATACATGCGCATGCGTCTCAGAACCTTA 992
Db 332 value---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIle 350
Qy 993 TACCTAGCTCTGAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1052
Db 351 TyrLysThrSerGlySerSerLeuAspTpsAlaTyrAspLeuGlyIleLysHisThrPhe 370
Qy 1053 AC-----ATCAAA 1060
Db 371 AlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLys 390
Qy 1061 CCCACCTGTAGAGAGCTTTTGGCGGTGCTCTAAATAGCTTGGCATGTCATTAGGAAT 1120
Db 391 ProThrCysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysHis 410

RESULT 12
QBN4T0 ID QBN4T0 PRELIMINARY; PRT; 437 AA.
AC QBN4T0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to carboxypeptidase A6.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033684; AAI33684.1; -.
DR Genew; HGNC:17245; CPA6.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR00834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase.
SQ SEQUENCE 437 AA; 51024 MW; 9970680D429A0A98 CRC64;

Alignment Scores:
Pred. No.: 8.36e-55 Length: 437
Score: 633.50 Matches: 147
Percent Similarity: 49.66% Conservative: 70
Best Local Similarity: 33.64% Mismatches: 147
Query Match: 22.81% Indels: 74
Gaps: 4

```

```

US-09-980-881A-1 (1-1573) x QBN4T0 (1-437)
Qy 9 GCTGTTGGATGAAGCTTTTC-----AGCCTTGCAGTCTCTGTACCCATTGTCTCTTC 62
Db 12 AlaAlaPheLeuProLeuCysTpsThrPheLeuLysIleLeuGlnPro----- 27
Qy 63 TGTGACAGCATGCTCTTC-----GGTTCAGAGTGGCCAGTTCTAGTCTGTCTCTCCT 116
Db 28 GlyHisSerHisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIlePro 47
Qy 117 AGAACCTCTAGGCAAGTCAAGTTCTACAGAATCTTACTACAACATATAGATGTTGTTCTC 176
Db 48 LysThrGluGluGluAlaTyrAlaLeuLysLysIleSerTyrGlnLeuLysValAspLeu 67
Qy 177 TGCAGCCCGGTAAACAGCTGACCTTATTGTGAAG-----AAAAAACAAGTCCATTTTTT 230
Db 68 TpsGlnProSerSerIleSerTyrValSerGluGlyThrValThrAspValHisIlePro 87
Qy 231 GTAAATGCATCTGATGTGCAATGTAAGCCCATTTAAATGTGAGCGGAATTCATGC 290
Db 88 GlnAsnGlySerArgAla-----LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyr 105
Qy 291 AGTGTCTTCTGTCAGACGCTGGAGATCTTATTCAACAGCAGATTTCCACAGCACAGTC 350
Db 106 LysValIleLeuGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGln 125
Qy 351 AGCCCGCCGACCTCCGCATCG-----TACTATGAACAGATCCTCCTCACTCACTCAAAATG 404
Db 126 ArgAsnArgArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluIle 145
Qy 405 TATTCTTGATAGATTTTAACTGAGAGCATCTCTGATATGTTTCAAAAATCCACATT 464
Db 146 GlnAsnTrpMetHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIle 165
Qy 465 CGATCTCTCATTTGAGAGTACCCACTCTATGTTTAAAGGTTTCTCGGAAAAGAACAAACA 524
Db 166 GlyArgSerTyrGluGlyArgCysLeuPheIleLeuLysLeu---GlyArgSerArg 184
Qy 525 GCCAAAATGCCATATGAGTTGACTGTGGAATCCATGCCAGAGATGAGTCTCTCTCTGCT 584
Db 185 LeuLysArgAlaValTrpIleAspCysGlyIleAlaArgGluTrpIleGlyProAla 204
Qy 585 TTCTGCTGTGTTTCATA----- 602
Db 205 PheCysGlnTrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArg 224
Qy 602 ----- 602
Db 225 LysMetLeuAsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHis 244
Qy 603 -----GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACATCAT 653
Db 245 PheSerTrpThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArg 264
Qy 654 TGCATCGGAACAGACCTGAAATAGCAACTTTGTCTCCAAAACACTGGTGTGAGGAGGTGCA 713
Db 265 CysArgGlyValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAla 283
Qy 714 TCAGTTCTCATGTCGGAAACCTACTGTGGACTTTATCTCTGAGTCAGAACCAAGAGTG 773
Db 284 SerMetHisProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluVal 303
Qy 774 AAGCAGTGGCTAGTTCTTGAGAGAAATATCAACAGATTAAGCATCATCAGCATG 833
Db 304 LysAlaValAlaAsnPheLeuArgLysHisIleArgAlaTyrLeuSerPhe 323
Qy 834 CATTTCATCTCCAGCATATAGTGTTCATATTTCTTATACAGAAAGTAAAGCAAGAC 893
Db 324 HisAlaTyrAlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsn 343
Qy 894 CATGAGGAACGTCTCTAGTAGCAGGTGAAGCAGGTGCTGCTATTGACAAACTAGTAAA 953

```

```

Db 344 PheArgCysValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyr 362
Qy 954 AATACCAAGGTATACACATGGCTGCAGAACCTTATACCTAGCTCTCGAGGTGGG 1013
Db 363 GlyValArgTyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySer 382
Qy 1014 GACCATGGATCTATGATTTGGGCATCAAAATATTCTGTTTAC----- 1054
Db 383 MetAspTrpAlaTyrLysAsnGlyLeProTyrAlaPheAlaPheGluLeuArgAspThr 402
Qy 1055 -----ATCAACCCACCTCTAGAGAGCTTTT 1081
Db 403 GlyTyrPheGlyPheLeuLeuProGluMetLeuLeuLysProThrCysThrGluThrMet 422
Qy 1082 GCCGCTCTCTAAATAGCTTGCATGCTCATAGGATGTTTAATGCCCC 1132
Db 423 LeuAlaValLysAsnIleThrMetHisLeuLeuLys-----CysPro 437

RESULT 13
Q8AVS2 PRELIMINARY; PRT; 434 AA.
AC Q8AVS2
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to carboxypeptidase A6.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
CX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC041315; AAH41315.1; -.
KW Carboxypeptidase.
SQ SEQUENCE 434 AA; 50581 MW; B42BD77843910591 CRC64;

Alignment Scores:
Pred. No.: 6.01e-54 Length: 434
Score: 625.00 Matches: 147
Percent Similarity: 49.30% Conservative: 65
Best Local Similarity: 34.19% Mismatches: 139
Query Match: 22.51% Indels: 80
DB: 13 Gaps: 11

US-09-980-881A-1 (1-1573) x Q8AVS2 (1-434)
Qy 39 GTCCTGTACCATGTTCTCTTCTGTGAGCAGCATGTTCTCGGTTCCAGAGTGGCCAA 98
Db 11 ValLeuLeuPheLeuLeuSerCys-----AlaLeuSerAlaGlyGln 25
Qy 99 -----GTTCTAGCTGCTCTCTCTAGAAC 122
Db 26 CysHisLeuTyrAsnAsnArgTyrSerTrpAspLysValIleArgValPheProLysAsn 45
Qy 123 TCTAGGCAAGTCAAGTTCTACAGAACTTACTACAATATGAGATTGTTCTCTGCGAG 182
Db 46 GluArgGluAlaAsnGluLeuLysAspMetTyrGlnGlnLeuGlnValAspLeuTrpGln 65
Qy 183 CCGTAACAGCTGACCTTATTGTAGNAA-----AAACAGTCCATTCTTTGTAAT 236
Db 66 ProSerSerIleSerHisIleGlyLysAspThrValThrAspValHisThrSerGlyAsn 85
Qy 237 GCATCTGATGTCGACAAATGCAAGCCCAATTAATGTCAGCGAAATTCATGTCAGTGC 296
Db 86 SerSerGluVal-----LeuLeuThrTyrLeuThrLysAlaLysIleGlnHisLysIle 103
Qy 297 TTGCTGCGACAGCTGGAGATCTTATTCAA---CAGCAGATTTCACACGACACAGTCAGC 353
Db 104 LeuValAsnAsnValGlnSerMetLeuGluAlaGlnAlaPheArgProArgArgLys 123

```

```

Qy 354 CCCGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCT 410
Db 124 ArgArgSerLeuSerArgTyrAsnTyrAsnGluTyrHisProLeuHisGluIleGluSer 143
Qy 411 TGGATAGAATTATTAATCTGAGAGGCATCTGATATGATCTTACAAAATCCACATTGTGATCC 470
Db 144 TrpMetPheTyrMetAsnLysThrHisAspLeuValSerLeuPheThrIleGlyLys 163
Qy 471 TCATTTGAGAAGTACCACCTCTATGTTTAAAGTTTCTGGAAAAGAAACAACAGCCAAA 530
Db 164 SerTyrGluGlyArgSerLeuTyrValLeuLysLeuGlyLysAspThrLysSerTyrLys 183
Qy 531 AATGCCATATGGATTGACCTGTGGAATCCATGCCAGAGATGATCTCTCTGCTTTCTGC 590
Db 184 LysAlaIleTrpIleAspCysGlyMetHisAlaArgGluTrpIleGlyProAlaPheCys 203
Qy 591 TTGTGTTTCATA----- 602
Db 204 GlnTrpPheValLysGluAlaIleAsnSerTyrAsnThrAspProAlaMetLysLysIle 223
Qy 603 -----GGCCAT----- 608
Db 224 LeuAsnLeuLeuTyrIleTyrValMetProValPheAsnValAspGlyTyrHisTyrSer 243
Qy 609 -----AATCGAATGTGGAGAAAGAACCGTTCTTCTATCGGAACAATCATTCATC 659
Db 244 TrpHisSerAspArgPheTrpArgLysThrArgSerLysAsnThrArgTyrGlnCysTyr 263
Qy 660 GGAACAGACCTGAATAGCACTTTGTCTCCAAACACTGGTGTGAGGAAGTGCATCCAGT 719
Db 264 GlyValAspAlaAsnArgAsnTrp---LysValHisTrpSerAspGluGlyAlaSerLeu 282
Qy 720 TCCTCATGTCTCGAAACCTACTGTGACATTATCTCTGAGTCAGAACCCAGAGTGAAGGCA 779
Db 283 AsnProCysAspAsnThrTyrCysGlyProTyrAlaGluSerGluProGluValLysAla 302
Qy 780 GTGCTAGTTTCTTGAGAGAAATATCAACACAGATTAAAGCATACATCAGCATGCTTCA 839
Db 303 ValAlaGlnPheLeuTyrLysGlnArgLysHisValArgAlaTyrMetSerPheHisAla 322
Qy 840 TACTCCGACATATAGTGTTCATATCTTATACAGAACTTATACAGAAAGAACCATGAG 899
Db 323 TyrAlaGlnMetLeuLeuTyrProTyrSerTyrGlnTyrGlyAlaIleProAsnPheGly 342
Qy 900 GAACCTGTCTAGTACCGAGTGAAGCAGTTCGTCTATTTCACAAAACACTAGTAAATACC 959
Db 343 CysValGluSerAlaAlaHisAsnAlaValLeuAlaIle---ArgSerAlaTyrGlyIle 361
Qy 960 AGGTATACACATGCCATGCTCAGAAACCTTATACCTAGCTCTCGAGTGGGACGAT 1019
Db 362 ArgTyrArgHisGlyProAlaSerSerThrLeuTyrLeuThrSerGlySerSerMetAsp 381
Qy 1020 TGGATCTATGATTGGGCATCAAAATATTCTGTTTAC----- 1054
Db 382 TrpAlaTyrAsnAsnGlyIleProTyrSerTyrAlaPheGluLeuArgAspThrGlyTyr 401
Qy 1055 -----ATCAACCCACCTGTAGAGAGCTTTTGGCGCT 1087
Db 402 TyrGlyPheLeuLeuProGluGlyLeuIleLysProThrCysValGluThrMetLeuAla 421
Qy 1088 GTCTCTAAATAGCTTGGCATGTCATTAGG 1117
Db 422 ValLysAsnIleThrMetHisAlaLeuLys 431

RESULT 14
Q8UUK1 PRELIMINARY; PRT; 419 AA.
AC Q8UUK1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preprocarboxypeptidase A precursor.

```

OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=domesticus; TISSUE=Pancreas;
 RA Hasegawa S., Honda K., Hikami Y.;
 RT "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A precursor."
 RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X64539; CAA45837.1; --
 DR MEROSES; M14.001; --
 DR InterPro; IPR001146; Propep M14.
 DR InterPro; IPR008334; Zn carboxypeptidase.
 DR Pfam; PF02244; Propep M14; 1.
 DR Pfam; PF00246; Zn carboxypeptidase; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR SMART; SM00631; Zn_pept; 1.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Signal; Carboxypeptidase.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 113 419 CARBOXYPEPTIDASE A.
 SQ SEQUENCE 419 AA; 47006 MW; 299C334274BCB9C CRC64;
 Alignment Scores:
 Pred. No.: 1.63e-49 Length: 419
 Score: 581.00 Matches: 133
 Percent Similarity: 52.52% Conservative: 65
 Best Local Similarity: 35.28% Mismatches: 107
 Query Match: 20.92% Indels: 72
 DB: 13 Gaps: 9
 US-09-980-881A-1 (1-1573) x Q8UUK1 (1-419)
 QY 84 TTCAGAGTGGCGAAGTTCTAGTCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
 DB 21 PheValGlyHisGlnValLeuArgIleValProSerSerAspAlaGluLeuGlnLysVal 40
 QY 144 CAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCGGTACAGCTGACCTATT 203
 DB 41 GlnGluLeuGlnGlu----- 45
 QY 204 GTGAAGAAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAAT----- 254
 DB 46 LeuGluHisLeuGlnLeuAspPheTrpLeuSerProArgGlyLeuGlyAsnProValAsp 65
 QY 255 -----GTGAAGCCCATTTAAATGTGACGCGAATT 284
 DB 66 IleArgValProPheProSerLeuGlnProValLysAlaHisLeuGluAlaAsnGlyVal 85
 QY 285 CCATGAGTCTCTGTCGGCAGAGTGGAGATCTTATT-----CAACAGCAG 332
 DB 86 ProTyrSerIleMetIleGluAspValGlnAlaLeuValAspLeuGluGlnMetGlnMet 105
 QY 333 ATTTCCAAACGACACAGTCAGCCCGAGCTCCGATCGTAC--TATGAACAGTATCAC 389
 DB 106 LeuArgArgArgArgPheValProLeuSerThrSerThrPheAspTyrThrSerTyrHis 125
 QY 390 TCACTAAATGAATCTATTCTTGGATAGAAATTTTAACTGAGAGGCTCTCTGATATGCTT 449
 DB 126 ThrLeuGluGluIleTyrAlaPheMetAspLeuLeuValAlaGluAsnProAsnLeuVal 145
 QY 450 ACAAATCCACATTCATCTCATTTGAGAAGTACCCTCTATGTTTAAAGTTTCT 509
 DB 146 SerLysLeuGluIleGlyArgThrThrGluAsnArgProIleTyrValLeuLysPheSer 165
 QY 510 GGAAGAACAAACAGCCAAATGCCATGATGACGTGCAATCCATGCCAGAGAA 569
 DB 166 ---LysGlyGlyThrAsnArgProAlaIleTrpIleAspThrGlyIleHisSerArgGlu 184

QY 570 TGGATCTCTCCCTGCTTTCTGCTGCTGCTCATA----- 602
 DB 185 TrpValThrGlnAlaSerGlyIleTrpPheAlaLysLysIleValGlnGluGlnAspGlu 204
 QY 602 ----- 602
 DB 205 GlyLeuAlaAsnIleLeuAspGlnMetAspIlePheLeuGluIleValThrAsnProAsp 224
 QY 603 -----GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCG 644
 DB 225 GlyPheAlaPheThrHisThrSerAsnArgMetTrpArgLysThrArgSerLysArgSer 244
 QY 645 ACAATCATTCATCGACACACACACCTGAATAGCAACTTGTCTCCAAACACCTGCTGTGAG 704
 DB 245 GlySerLeuCysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGly 263
 QY 705 GAAGGTGCATCCAGTTCCTCATCTCGGAAACCTACTGTGGACTTTTATCCTGAGTCAGAA 764
 DB 264 SerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyProTyrAlaAsnSerGlu 283
 QY 765 CCAGAAAGTGAAGCAGTGGCTAGTTTCTTGGAGAGAAATATCAACAGATTTAAAGCATAC 824
 DB 284 ProGluValLysAlaIleValAspPheValLysSerHisLysAsn---IleLysAlaPhe 302
 QY 825 ATCAGATGCATTCATCTCCAGCATATAGTGTTCATATTCCTATATACACAGAAATAA 884
 DB 303 IleSerIleHisSerTyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrAla 322
 QY 885 AGCAAGACCATCAGGAACTGCTCTAGTAGCCAGTGAAGCAGTTCGTCGATTGACAA 944
 DB 323 ValProAspLysGluLeuHisGlnValAlaLysGluAlaValAlaLeu---Ser 341
 QY 945 ACTAGTAAATAATACAGGTATACACATGGCATGGCTCAGAAACCTTATACCTAGCTCCT 1004
 DB 342 SerLeuTyrGlyThrAsnTyrLysTyrGlySerIleIleThrThrIleTyrGlnAlaSer 361
 QY 1005 GGAGTGGGGCAGATGGATCTATGATTTGGGCATCAAAATATTCGTTTACA 1055
 DB 362 GlyGlyThrIleAspTrpThrTyrAsnGlnGlyIleLysTyrSerPheThr 378
 RESULT 15
 Q8IVL8 PRELIMINARY; PRT; 374 AA.
 AC Q8IVL8;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Zn-carboxypeptidase.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RA Obaya A.J., Lopez-Otin C.;
 RT "A new Zn-carboxypeptidase highly expressed in ovary."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ422118; CAD19478.1; --
 KW Carboxypeptidase.
 SQ SEQUENCE 374 AA; 42529 MW; 404C373BB841AAD2 CRC64;
 Alignment Scores:
 Pred. No.: 2.04e-48 Length: 374
 Score: 570.00 Matches: 117
 Percent Similarity: 52.47% Conservative: 53
 Best Local Similarity: 36.11% Mismatches: 90
 Query Match: 20.53% Indels: 65
 DB: 4 Gaps: 7
 US-09-980-881A-1 (1-1573) x Q8IVL8 (1-374)
 QY 324 CAACAGCATTTTCCAAACGACACAGTCAGCCCCCGAGCCTCCGTCATCGTAC---TATGAA 380


```
Db 29 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn 48
Qy 381 CAGTATCACTCACTAAATCAATCTATCTTGGATAGAAATTTATACTGAGAGCATCCT 440
Db 49 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 68
Qy 441 GATATGCTTACAAAAATCCACATGGATCCCTCATTTGAGAAGTACCCCACTCTATGTTTAA 500
Db 69 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProMetTyrTyrLeu 88
Qy 501 AAGGTTTCTGAAAGAACCAACCAACCAAAATATCCATATGATGATGATGATGATGATGAT 560
Db 89 LysIleSerGlnProSerGlyAsnProLysIleIleIleTrpMetAspCysGlyIleHis 108
Qy 561 GCCAGAGATGGATCTCTCTGCTTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 602
Db 109 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 128
Qy 602 ----- 602
Db 129 HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 148
Qy 603 -----GCCCAT-----AATCGAATGCTGGAAGAAAGAAC 629
Db 149 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer 168
Qy 630 CGTTCTTCTTATGCGAATCAATCATTTGTCATCGAAGACAGACCTGAATAGCACTTTGTCTCC 689
Db 169 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 187
Qy 690 AAACACTGCTGTGAGGAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTT 749
Db 188 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 207
Qy 750 TATCCTGAGTCAGAACAGAGTGAAGCAGTCAGTCTAGTTTCTTGAAGAAATATCAAC 809
Db 208 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 227
Qy 810 CAGATTAAAGCATACATCAGCATCATCATCTCCAGCATATAGTGTGTTTCCATATTC 869
Db 228 AspileLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 247
Qy 870 TATACGAGTAAAGCAAGACATCAGCACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
Db 248 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 267
Qy 930 CGTGTATTGACAAAACTAGTAAATAATACAGGTATATACATGCGCCATGCTCAGAAACC 989
Db 268 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspile 286
Qy 990 TTATACCTAGCTCTGAGGTGGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1049
Db 287 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspileGlyIleProPheSer 306
Qy 1050 TTTACATC----- 1057
Db 307 TyrThrPheGluLeuArgAspSerGlyThrTyrGlyPheValLeuProGluAlaGlnIle 326
Qy 1058 AAACCCACCTGTAGAGAGCTTTTGGCGCTGCTCTCTAAATAGCT----- 1102
Db 327 GlnProThrCysGluGluThrMetGluAlaValLeuSerValLeuAspAspValTyrAla 346
Qy 1103 -----TGGCAT 1108
Db 347 LysHisTrpHis 350
```

Search completed: January 1, 2004, 18:58:33
Job time : 106.358 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:58:24 ; Search time 5920.84 Seconds
(without alignments)
10868.540 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattctgttggtg9.....aaaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_cm.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1528	97.1	1546	9	AB011969	AB011969 Homo sapi
2	1469.8	93.4	1625	6	AR232183	AR232183 Sequence
3	1372	87.2	1728	6	AX409750	AX409750 Sequence
4	1372	87.2	1728	6	AX706330	AX706330 Sequence
5	1372	87.2	1728	9	HUMPCPBX	M75106 Human prepr
6	1372	87.2	1749	6	I16100	I16100 Sequence 2
7	1372	87.2	1749	6	I33526	I33526 Sequence 2
8	1309	83.2	1715	9	BC007057	BC007057 Homo sapi
9	918	58.4	1272	6	AR086324	AR086324 Sequence
10	915.4	58.2	1272	6	BD106653	BD106653 Method of
11	914.4	58.1	1272	9	BT006936	BT006936 Homo sapi
12	914.4	58.1	1272	12	BT007990	BT007990 Synthetic
13	656	41.7	1421	10	AB021968	AB021968 Mus muscu
14	656	41.7	1490	10	AF164524	AF164524 Mus muscu
15	654	41.6	1430	10	AF186188	AF186188 Mus muscu
16	652.6	41.5	1425	10	AB042598	AB042598 Rattus no
17	478.6	30.4	55827	6	AR232184	AR232184 Sequence
18	475.4	30.2	58097	9	AL157758	AL157758 Human DNA
19	409	26.0	437	11	G05971	G05971 human STS w
20	173.6	11.0	181	11	G59992	G59992 SHGC-130613
21	156	9.9	251806	2	AC096305	AC096305 Rattus no
22	156	9.9	258265	2	AC120472	AC120472 Rattus no
23	137.2	8.7	431	6	AX400792	AX400792 Sequence
24	128	8.1	180680	9	AL137141	AL137141 Human DNA
25	124	7.9	927	6	AR072911	AR072911 Sequence
26	123.8	7.9	1327	5	AF190274	AF190274 Bothrops
27	123.4	7.8	921	6	I67699	I67699 Sequence 3
28	123.4	7.8	1215	6	I67698	I67698 Sequence 1
29	123.4	7.8	1251	4	SSC133775	AJ133775 Sus scrof
30	123.2	7.7	451	11	G53326	G53326 SHGC-82320
31	120.8	7.7	1251	4	D0G2AP47	D78348 Canis famli
32	109	6.9	1263	6	AB7530	AB7530 Sequence 11
33	107.6	6.8	1622	9	HUMCARMC	M27717 Human mas
34	107.6	6.8	1622	11	G28614	G28614 human STS S
35	107.6	6.8	1674	9	BC012613	BC012613 Homo sapi
36	107.4	6.8	999	6	A51908	A51908 Sequence 72
37	107.4	6.8	999	6	AR085871	AR085871 Sequence
38	107.4	6.8	1053	6	A51904	A51904 Sequence 68
39	107.4	6.8	1053	6	AR085868	AR085868 Sequence
40	107.4	6.8	1254	12	BT008111	BT008111 Synthetic
41	107.4	6.8	1263	6	A51896	A51896 Sequence 60
42	107.4	6.8	1263	6	AR085861	AR085861 Sequence
43	107.4	6.8	1284	6	A51913	A51913 Sequence 77
44	107.4	6.8	1284	6	AR085875	AR085875 Sequence
45	107.4	6.8	1310	9	HSA224866	AJ224866 Homo sapi

ALIGNMENTS

RESULT 1
AB011969
LOCUS
DEFINITION Homo sapiens mRNA for carboxypeptidase B-like protein, complete cds.
1546 bp mRNA linear PRI 02-FEB-2000

ACCESSION AB011969
VERSION AB011969.1 GI:6855463
KEYWORDS Carboxypeptidase B-like protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
REFERENCE
AUTHORS Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE Isolation, molecular cloning, and partial characterization of a

novel carboxypeptidase B from human plasma
J. Biol. Chem. 266 (32), 21833-21838 (1991)
92042093
1939207
2 (sites)
Matsumoto, A.
isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
Unpublished
3 (bases 1 to 1546)
Matsumoto, A.
Direct Submission
Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
Medicine, Department of Radiation Biophysics & Genetics,
Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
(E-mail: amate@med.kobe-u.ac.jp, Tel: 81-78341-7451)
Sequence updated (10-Apr-1998).
Location/Qualifiers
1..1546
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
20..1102
/note="carboxypeptidase B-like protein in human brain"
/codon_start=1
/product="carboxypeptidase B-like protein"
/protein_id="BA90475.1"
/db_xref="GI:6855464"
/translations="MKLCSLAVLPVILFCEHVFAPQTQGVLAALPRTSRQVQLQN
LTTTIVLPQPTADLIKKQVHFVNASVDVNVKAHLNLSGI PCSVLLADVEDLI
QQQISNDTSPRASASYEQHSEINLWIEFIETHPDMLTKIHIGSFEKYPVLY
LKYSKGEOKAKNAIWDGCHAREWISPAFLWFIHNRMRKNRSFYANNHCIGIDL
NRNPSKHWCEEGASSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYS
OHVFPYSYTRSKDHELSVASEAVRAIEKISKNTRYTHGSGSETLYLAPGGDD
WIYDLGKVSFTSNPPVEKLPLSLK"
1546
polyA_site
/note="21 A nucleotides"
BASE COUNT 453 a 323 c 297 g 473 t
ORIGIN
Query Match 97.1%; Score 1528; DB 9; Length 1546;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1534; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 AGAAATTCGTGGGATGAAGCTTTGACGCTTGACGCTTGTACCCATTGTTCTCT 60
DB 3 AGAAATTCGTGGGATGAAGCTTTGACGCTTGACGCTTGTACCCATTGTTCTCT 62
QY 61 TCTGTAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAA 120
DB 63 TCTGTAGCAGCATGTCTTCGCGTTCCAGACTGGCCAAAGTTCTAGCTGCTCTTCTAGAA 122
QY 121 CTTCTAGGCAAGTTCAGTTCACAGAACTTACTACAAATATGAGATTGTTCTCTGGC 180
DB 123 CTTCTAGGCAAGTTCAGTTCACAGAACTTACTACAAATATGAGATTGTTCTCTGGC 182
QY 181 AGCCGGTAAAGCTGACCTTATGTGAGAGAAAACAAAGTCCATTTTGTAAATGCAT 240
DB 183 AGCCGGTAAAGCTGACCTTATGTGAGAGAAAACAAAGTCCATTTTGTAAATGCAT 242
QY 241 CTGATGTCGACAAATGTCAAGAGCCCATTTAAATGTGAGCGAAATTCATGCAAGTGTCTGC 300
DB 243 CTGATGTCGACAAATGTCAAGAGCCCATTTAAATGTGAGCGAAATTCATGCAAGTGTCTGC 302
QY 301 TGGCAGACGTGGAGATCTTATTCAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAG 360
DB 303 TGGCAGATGTGGAAGATCTTATTCAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAG 362
QY 361 CCTCCGATCGTACTATGACAGTATCACTCACTAAATGAATCTATTCTTGATAGAAAT 420
DB 363 CCTCCGATCGTACTATGACAGTATCACTCACTAAATGAATCTATTCTTGATAGAAAT 422
QY 421 TTATAACTGAGAGCATCCTGATATGCTTACAAAATCCACATTTGGATCCTCATTTGAGA 480
|||||
423 TTATAACTGAGAGCATCCTGATATGCTTACAAAATCCACATTTGGATCCTCATTTGAGA 482
QY 481 AGTACCCACTATATGTTTTAAAGTTTTCTGGAAGAAACAAACAGCCAAAATATGCCATAT 540
DB 483 AGTACCCACTATATGTTTTAAAGTTTTCTGGAAGAAACAAACAGCCAAAATATGCCATAT 542
QY 541 GGATTCACCTGTGAATCCATGCGAGAGATGGAATCTCTCTGCTTCTGCTTCTGCTGTTCA 600
DB 543 GGATTCACCTGTGAATTCATGCGAGAGATGGAATCTCTCTGCTTCTGCTTCTGCTGTTCA 602
QY 601 TAGGCCATATTCGAATGTGGAAGAAACCGTTCTTCTTATGCGAACAATCATTTGCATCG 660
DB 603 TAGGCCATATTCGAATGTGGAAGAAACCGTTCTTCTTATGCGAACAATCATTTGCATCG 662
QY 661 GAACAGACCTGAATAGCAACTTTGCTCTCAACACACTGGTGTGAGGAGGTGCATCCAGTT 720
DB 663 GAACAGACCTGAATAGCAACTTTGCTCTCAACACACTGGTGTGAGGAGGTGCATCCAGTT 722
QY 721 CCTCATGCTCGGAACCTACTCTGCGACTTTATCTCTGAGTCAGAACCCAGAGTCAAGGCAG 780
DB 723 CCTCATGCTCGGAACCTACTCTGCGACTTTATCTCTGAGTCAGAACCCAGAGTCAAGGCAG 782
QY 781 TGCTTAGTTCTTGTGAGAGAAATATCAACAGATTAAGACATACATCAGCATCATTTTAT 840
DB 783 TGCTTAGTTCTTGTGAGAGAAATATCAACAGATTAAGACATACATCAGCATCATTTAT 842
QY 841 ACTCCAGCATATAGTGTGTTTCCATATTCCTATACAGAAAGTAAAGAACATCAGTGG 900
DB 843 ACTCCAGCATATAGTGTGTTTCCATATTCCTATACAGAAAGTAAAGAACATCAGTGG 902
QY 901 AACTGTCTCTAGTAGCAGTGAAGCAGTTCGCTCTATTCACAAAACCTAGTAAAAATACCA 960
DB 903 AACTGTCTCTAGTAGCAGTGAAGCAGTTCGCTCTATTCAGAAAATTTAGTAAAAATACCA 962
QY 961 GGTATACATGCGCCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCAT 1020
DB 963 GGTATACATGCGCCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCAT 1022
QY 1021 GGATCTATGATTTGGGATCAAAATATTCGTTTACATCAACCCACCTCTAGAGAGCTTT 1080
DB 1023 GGATCTATGATTTGGGATCAAAATATTCGTTTACATCAACCCACCTCTAGAGAGCTTT 1082
QY 1081 TGGCGGTCTCTAAATAGCTTGGCATGTCTTAGGAATGTTTAAATGCCCTGATTTTA 1140
DB 1083 TGGCGGTCTCTAAATAGCTTGGCATGTCTTAGGAATGTTTAAATGCCCTGATTTTA 1142
QY 1141 TCATCTGCTTCGCTATTTTAAATTTACTGATTCAGCAAGACCAAAATCATTTGTATCAGAT 1200
DB 1143 TCATCTGCTTCGCTATTTTAAATTTACTGATTCAGCAAGACCAAAATCATTTGTATCAGAT 1202
QY 1201 TATTTTAAAGTTTTATCCGATGTTTTGATAAAGATTTTCTATTCCTTGGTGTCTGTCAG 1260
DB 1203 TATTTTAAAGTTTTATCCGATGTTTTGATAAAGATTTTCTATTCCTTGGTGTCTGTCAG 1262
QY 1261 AGAACCTTAATAGTGTCTTTGCCATTAAGGAGACTAGGGTTCATGCTTTTACCCCT 1320
DB 1263 AGAACCTTAATAGTGTCTTTGCCATTAAGGAGACTAGGGTTCATGCTTTTACCCCT 1322
QY 1321 TTAATAAAAAATTTGTAAGAGTCTAGTATCCTACTTTTCTTGTGATTTTCGACGTTTGACT 1380
DB 1323 TTAATAAAAAATTTGTAAGAGTCTAGTATCCTACTTTTCTTGTGATTTTCGACGTTTGACT 1382
QY 1381 AGCCATCTCAAGCAACTTTTCGACGTTTGATAGCCATCTCAAGCAAGTTTTAAATCAAGAT 1440
DB 1383 AGCCATCTCAAGCAACTTTTCGACGTTTGATAGCCATCTCAAGCAAGTTTTAAATCAAGAT 1442
QY 1441 CATCTCACGCTGATCATTTGGATCCTACTCAACAAAAGGAGGGTGTGAGAGTACATTA 1500
DB 1443 CATCTCACGCTGATCATTTGGATCCTACTCAACAAAAGGAGGGTGTGAGAGTACATTA 1502
QY 1501 AAGATTTCTGCTCAAAATTTTCAATAAATTTCTTTCTCTCTTT 1544
|||||

[illegible]

RESULT 3			
AX409750			
LOCUS	AX409750	1728 bp	DNA
DEFINITION	Sequence 2397 from Patent WO0225103.		
ACCESSION	AX409750	linear	PAT 14-JUN-2002

VERSION	AX409750.1	GI:21442455	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.		
AUTHORS	Gene expression profiles in liver cancer		
TITLE	Patent: WO 0229103-A 2397 11-APR-2002;		
JOURNAL	GENE LOGIC INC (US)		
FEATURES	Location/Qualifiers		
source	1..1728		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/notes="EMBL/GenBank Accession No. M75106"		
BASE COUNT	518 a 354 c 338 g 518 t		
ORIGIN			
Query Match	87.28; Score 1372; DB 6; Length 1728;		
Best Local Similarity	90.34; Pred. No. 2.6e-306;		
Matches 1558; Conservative	0; Mismatches 5; Indels 163; Gaps 2;		
Qy	1	AGAAATTTGCTGTGGGATGAAGCTTTGCGGTTCCAGATGCTTGTACCCATTGTTCTCT	60
Db	3	AGAAATTTGCTGTGGGATGAAGCTTTGCGGTTCCAGATGCTTGTACCCATTGTTCTCT	62
Qy	61	TCGTGAGCAGCATGTCTTCCGGTTCCAGATGCTTGTACCCATTGTTCTCTCTAGAA	120
Db	63	TCGTGAGCAGCATGTCTTCCGGTTCCAGATGCTTGTACCCATTGTTCTCTCTAGAA	122
Qy	121	CCTCTAGGCAAGTTCAGATGCTTCTACAGATCTTCTACCAATATGATGATGTTCTCTGC	180
Db	123	CCTCTAGGCAAGTTCAGATGCTTCTACAGATCTTCTACCAATATGATGATGTTCTCTGC	182
Qy	181	AGCGGTAAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT	240
Db	183	AGCGGTAAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT	242
Qy	241	CTGATGTCGAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTGC	300
Db	243	CTGATGTCGAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTGC	302
Qy	301	TGCGAGAGTGGAGATCTTATTAACAGCAGATTTTCCACGACACAGTCCAGCCCGAG	360
Db	303	TGCGAGAGTGGAGATCTTATTAACAGCAGATTTTCCACGACACAGTCCAGCCCGAG	362
Qy	361	CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATGAAT	420
Db	363	CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATGAAT	422
Qy	421	TTATACTGAGAGCATCTGATATGCTTACAAATATCCATTTGGATCTCTCAATTTGAGA	480
Db	423	TTATACTGAGAGCATCTGATATGCTTACAAATATCCATTTGGATCTCTCAATTTGAGA	482
Qy	481	AGTACCCACTCTATGTTTAAAGTTTCTGAAAGAAACAAACAGCCAAAATGCCATAT	540
Db	483	AGTACCCACTCTATGTTTAAAGTTTCTGAAAGAAACAAACAGCCAAAATGCCATAT	542
Qy	541	GGATTGACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTGTGGTTCA	600
Db	543	GGATTGACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTGTGGTTCA	602
Qy	601	TAGGCCAT-----A	608
Db	603	TAGGCCATATAAATCAATTTCTATGGGATAATAGGGCAATATACCAATCTCTCTGAGGCTG	662
Qy	609	-----A	609
Db	663	TGGATTTCTATGTTATCCCGGTGGTAAATGTGACGGTATGACTACTATCGGAAAGA	722
Qy	610	ATCGAATGTGGAAGAAACCGTTCTTCTTCTATGCGAACAATCATTGTCATCGGAACAGC	669
Db	723	ATCGAATGTGGAAGAAACCGTTCTTCTTCTATGCGAACAATCATTGTCATCGGAACAGC	782
Qy	670	TGATAGCAACTTTGTTCTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATGCT	729
Db	783	TGATAGCAACTTTGTTCTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATGCT	842
Qy	730	CGGAAACCTTACTGTGACCTTTATCCCTGAGTCAGAACAGAGTGAAGGAGTGGCTAGTT	789
Db	843	CGGAAACCTTACTGTGACCTTTATCCCTGAGTCAGAACAGAGTGAAGGAGTGGCTAGTT	902
Qy	790	TCCTGAGAAGAAATATCAACAGATTAAAGCATATCATGATGATGATTCATATCCACGC	849
Db	903	TCCTGAGAAGAAATATCAACAGATTAAAGCATATCATGATGATGATTCATATCCACGC	962
Qy	850	ATATAGTGTTCATATTCCTATACAGGATTAAGCAAGCAAGCAAGTGAAGGAGTGGCTC	909
Db	963	ATATAGTGTTCATATTCCTATACAGGATTAAGCAAGCAAGCAAGTGAAGGAGTGGCTC	1022
Qy	910	TAGTAGCCAGTGAAGCAGTTGCTGCTATTGACAAAACTAGTAAAAATACCAAGTATACAC	969
Db	1023	TAGTAGCCAGTGAAGCAGTTGCTGCTATTGAGAAAACTAGTAAAAATACCAAGTATACAC	1082
Qy	970	ATGSCCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGGACGATTTGGATCTATG	1029
Db	1083	ATGSCCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGGACGATTTGGATCTATG	1142
Qy	1030	ATTGSGGCATCAAAATTTGCT-----	1050
Db	1143	ATTGSGGCATCAAAATTTGCTTTTCAATTGAATCTCGAGATACGGGCACATACGGATTCT	1202
Qy	1051	-----TTACATCAAAACCCACTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA	1097
Db	1203	TGCTGCGGAGCGTTACATCAAAACCCACTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA	1262
Qy	1098	TAGCTTGGCATGTTCATTAGGAATGTTTAAATGCCCTGATTTTATCATTTCTGCTCCGAT	1157
Db	1263	TAGCTTGGCATGTTCATTAGGAATGTTTAAATGCCCTGATTTTATCATTTCTGCTCCGAT	1322
Qy	1158	TTTAAATTTACTGATTCAGCAAGACCAATCATTTGATCATGATTTTAAAGTTTATC	1217
Db	1323	TTTAAATTTACTGATTCAGCAAGACCAATCATTTGATCATGATTTTAAAGTTTATC	1382
Qy	1218	CGTAGTTTGTATAAAGATTTTCTTATCTCTGCTGTTCTGTACAGAACCTTAAATAGTCT	1277
Db	1383	CGTAGTTTGTATAAAGATTTTCTTATCTCTGCTGTTCTGTACAGAACCTTAAATAGTCT	1442
Qy	1278	ACTTTGCCATTAAGGCAGACTAGGGTTTCATGCTTTTACCCCTTTAAAAAAAATTTGTA	1337
Db	1443	ACTTTGCCATTAAGGCAGACTAGGGTTTCATGCTTTTACCCCTTTAAAAAAAATTTGTA	1502
Qy	1338	AGCTAGTTTACTGATCTTTTCTGATTTTTCAGCTTTGATGCTAGCCATCTCAAGCACT	1397
Db	1503	AGCTAGTTTACTGATCTTTTCTGATTTTTCAGCTTTGATGCTAGCCATCTCAAGCACT	1562
Qy	1398	TTGAGTTTGTAGTACCATCTCAAGCAAGTTTTAAATCAAGATCATCTCAAGCTGATCAT	1457
Db	1563	TTGAGTTTGTAGTACCATCTCAAGCAAGTTTTAAATCAAGATCATCTCAAGCTGATCAT	1522
Qy	1458	TGATCTCTACTCAAAAAAGGAGGTGGTGCAGAGTACATTTAAAGATTTTCTGCTCCAA	1517
Db	1623	TGATCTCTACTCAAAAAAGGAGGTGGTGCAGAGTACATTTAAAGATTTTCTGCTCCAA	1682
Qy	1518	TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAAATTTTAAAAA 1563	
Db	1683	TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAAATTTTAAAAA 1728	
RESULT 4			
AX706330			
LOCUS	AX706330	1728 bp	DNA linear PAT 04-APR-2003
DEFINITION	Sequence 1 from Patent WO03014391.		
ACCESSION	AX706330		

AX706330.1	GI:29562754		
VERSION			
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE			
AUTHORS	Morten, J.E. and Jackson, K.N.		
TITLE	Method		
JOURNAL	Patent: WO 03014391-A 1 20-FEB-2003;		
	AstraZeneca AB (SE)		
FEATURES			
source	Location/Qualifiers		
	1..1728		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	518 a 354 c 338 g 518 t		
ORIGIN			
	Query Match 87.2%; Score 1372; DB 6; Length 1728;		
	Best Local Similarity 90.3%; Pred. No. 2.6e-306;		
	Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;		
Qy	1	AGAAATTGCTGTGGATGAAGCTTTGCAGCCTTGCAGTCTTGTACCCATTGTTCTCT	60
Db	3	AGAAATTGCTGTGGATGAAGCTTTGCAGCCTTGCAGTCTTGTATCCCAATTGTTCTCT	62
Qy	61	TCGTGAGCAGCATGCTTCGCGTTCCAGAGTGCCAAAGTTCTAGCTGCTCTTCTTAGAA	120
Db	63	TCGTGAGCAGCATGCTTCGCGTTCCAGAGTGCCAAAGTTCTAGCTGCTCTTCTTAGAA	122
Qy	121	CTCTAGGCAAGTTCAGTCTTACAGAACTTTACTACAAATATGAGATGTTCTCTGGC	180
Db	123	CTCTAGGCAAGTTCAGTCTTACAGAACTTTACTACAAATATGAGATGTTCTCTGGC	182
Qy	181	AGCGGTAAACAGCTGACCTTATTGTGAAGAAAACAAAGTCCATTTTGTAAATGCAT	240
Db	183	AGCGGTAAACAGCTGACCTTATTGTGAAGAAAACAAAGTCCATTTTGTAAATGCAT	242
Qy	241	CTGATGTGCACAAATGTGAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTTGC	300
Db	243	CTGATGTGCACAAATGTGAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTTGC	302
Qy	301	TGCGACAGTGGAAAGATCTTTATTCAACAGCAGATTTCCAAACGACACAGTCCGCGAG	360
Db	303	TGCGACAGTGGAAAGATCTTTATTCAACAGCAGATTTCCAAACGACACAGTCCGCGAG	362
Qy	361	CCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA	420
Db	363	CCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA	422
Qy	421	TTATAACTGAGAGGCATCTGATATGCTTTACAAAATCCACATTTGGATCTCTATTGAGA	480
Db	423	TTATAACTGAGAGGCATCTGATATGCTTTACAAAATCCACATTTGGATCTCTATTGAGA	482
Qy	481	AGTACCCACTCTATGTTTTAAAGTTTTCTGGAAGAAAACAAACAGCCAAAATGCCATAT	540
Db	483	AGTACCCACTCTATGTTTTAAAGTTTTCTGGAAGAAAACAAACAGCCAAAATGCCATAT	542
Qy	541	GGATTGACTGTGGAATCCATGCCAGAGATTTGATCTCTCTGCTTTCTGCTTGTGTTCA	600
Db	543	GGATTGACTGTGGAATCCATGCCAGAGATTTGATCTCTCTGCTTTCTGCTTGTGTTCA	602
Qy	601	TAGGCCAT-----	608
Db	603	TAGGCCATATACTCAATTTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTG	662
Qy	609	-----A	609
Db	663	TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAGAAGA	722
Qy	610	ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAAACAATCAATTCGATCGGAACAGACC	669

Db	723	ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAAACAATCAATTCATTCGATCGGAACAGACC	782
Qy	670	TGAATAGCAAACTTTGTCTCCAAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCT	729
Db	783	TGAATAGGAACCTTTGTCTTCCAAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCT	842
Qy	730	CGAAACCTTACTGTGGACTTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGACAGTGGCTAGTT	789
Db	843	CGAAACCTTACTGTGGACTTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGACAGTGGCTAGTT	902
Qy	790	TCCTTGAGAGAAATATCAACAGATTTAAAGCATATCATCAGCATGCAATCTACTTCCAGC	849
Db	903	TCCTTGAGAGAAATATCAACAGATTTAAAGCATATCATCAGCATGCAATCTACTTCCAGC	962
Qy	850	ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATAGAGAACTGTCTC	909
Db	963	ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATAGAGAACTGTCTC	1022
Qy	910	TAGTAGCCAGTGAAGCAGTTCTGCTATTTCACAAAACCTAGTAAATAACAGGTATATAC	969
Db	1023	TAGTAGCCAGTGAAGCAGTTCTGCTATTTCAGAAAACCTAGTAAATAACAGGTATATAC	1082
Qy	970	ATGCGCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGGATTTGGATCTATG	1029
Db	1083	ATGCGCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGGATTTGGATCTATG	1142
Qy	1030	ATTGGGCATCAAAATATTCGT-----	1050
Db	1143	ATTGGGCATCAAAATATTCGTTCACAACTTCGAGATACGGGCACATACGGATTCT	1202
Qy	1051	-----TTACATCAAAACCCCTGTAGAGAGCTTTTGGCGCTGTCTCTAAAA	1097
Db	1203	TGCTGCGGAGCGTTTACATCAAAACCCCTGTAGAGAGCTTTTGGCGCTGTCTCTAAAA	1262
Qy	1098	TAGCTTGGCATGTCATTAGGAATGTTTAAATGCCCTGATTTTATCATCTCTGCTCCGTAT	1157
Db	1263	TAGCTTGGCATGTCATTAGGAATGTTTAAATGCCCTGATTTTATCATCTCTGCTCCGTAT	1322
Qy	1158	TTTAAATTTACTGATTCAGCAAGACCAAAATCATTTGATCAGATTTATTTTAAAGTTTATC	1217
Db	1323	TTTAAATTTACTGATTCAGCAAGACCAAAATCATTTGATCAGATTTATTTTAAAGTTTATC	1382
Qy	1218	CGTAGTTTGTATAAAAGATTTTCTATTCCTTGGTCTGTGAGAGAACCTTAATAAGTCT	1277
Db	1383	CGTAGTTTGTATAAAAGATTTTCTATTCCTTGGTCTGTGAGAGAACCTTAATAAGTCT	1442
Qy	1278	ACTTTGCCATTAAGCAGACTAGGTTTCATGCTTTTTACCCCTTAAATAAATAATTTGTA	1337
Db	1443	ACTTTGCCATTAAGCAGACTAGGTTTCATGCTTTTTACCCCTTAAATAAATAATTTGTA	1502
Qy	1338	AACTCTAGTTTACTCTACTTTTCTTTGATTTTCGACGTTTTCAGCTAGCCATCTCAAGCACT	1397
Db	1503	AACTCTAGTTTACTCTACTTTTCTTTGATTTTCGACGTTTTCAGCTAGCCATCTCAAGCACT	1562
Qy	1398	TTCCAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATAAGATCATCTCTACGCTGATCAT	1457
Db	1563	TTCCAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATAAGATCATCTCTACGCTGATCAT	1622
Qy	1458	TGATCTCTACTCAAAAAGAGGTTGTCAGAGTACATTTAAAGATTTCTGCTCCAAA	1517
Db	1623	TGATCTCTACTCAAAAAGAGGTTGTCAGAGTACATTTAAAGATTTCTGCTCCAAA	1682
Qy	1518	TTTTTCAATAAATTTCTTCTTCTCTTTTAAAAAATAAATAAATAAATAAATAAATAA	1563
Db	1683	TTTTTCAATAAATTTCTTCTTCTCTTTTAAAAAATAAATAAATAAATAAATAAATAA	1728

RESULT 5	
HUMPCPBX	
LOCUS	
DEFINITION	Human prepro-plasma carboxypeptidase B mRNA, linear PRI 07-JAN-1995
ACCESSION	M75106
VERSION	M75106.1 GI:189686

```
KEYWORDS      plasma carboxypeptidase.
SOURCE        Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1728)
AUTHORS      Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE        Isolation, molecular cloning, and partial characterization of a
              novel carboxypeptidase B from human plasma
JOURNAL       J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE       92042093
COMMENT       PubMed 1939207
FEATURES      Original
              source text: Homo sapiens liver cDNA to mRNA.
              Location/Qualifiers
                1..1728
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /tissue_type="liver"
                1..1728
                /gene="pCPB"
                20..1291
                /gene="pCPB"
                /note="prepro-plasma carboxypeptidase B"
                /codon_start=1
                /protein_id="AA060042.1"
                /db_xref="GI:189687"
                /translation="MKLCSLAVLPVLFCEQHVFAFQSQVLAALPRTSRQVQLQN
                LTTYEILVWPQVADLIKKQVHFFVNASDVNKKALHVNKIPCSVLLADVBDLI
                QQISIDVSPRASAYEQYSLNIYSWIFETIHPDMLTKIHIGSFERYPLVY
                LKVGKEQAKNAIWDICGIHAREWISPAFLWFIQTQFYGIIGYINLLRLVDFY
                VMPVNVGDYDYSWKNRMRKNRSFYANNHCIGTDLNRNFASKHWCCEGASSCSSE
                TYGLYPESEPEKAVASFLRNINQIKAVISMHSYQHLVFPYSYTRSKDHEBLS
                LVASEAVRALEKSKNTRYTHGSETLYLAPGGDDMIYDLGIKISFTIELRDTGTI
                GFULPERYIKTCREAPAAVSKIAHWIRNV"
                86..1288
                /gene="pCPB"
                /product="plasma carboxypeptidase B"
                518 a 354 c 338 g 518 t
                mat_peptide
                BASE COUNT 518 a 354 c 338 g 518 t
                ORIGIN
                Query Match 87.2%; Score 1372; DB 9; Length 1728;
                Best Local Similarity 90.3%; Pred. No. 2.6e-306;
                Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
                QY 1 AGAAATTCGTGTGGGATGAAGCTTTCGAGCCTTGAGTCCTGTACCCATTGTTCTCT 60
                DB 3 AGAAATTCGTGTGGGATGAAGCTTTCGAGCCTTGAGTCCTGTACCCATTGTTCTCT 62
                QY 61 TCTGTGAGCAGCATGCTTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTTCCTAGAA 120
                DB 63 TCTGTGAGCAGCATGCTTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTTCCTAGAA 122
                QY 121 CCTCTAGGCAAGTTCAAGTTCTACAGAACTCTACTACAAATATGAGATTGTTCTCTGGC 180
                DB 123 CCTCTAGGCAAGTTCAAGTTCTACAGAACTCTACTACAAATATGAGATTGTTCTCTGGC 182
                QY 181 AGCCGGTAACAGCTGACCTTATTTGGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
                DB 183 AGCCGGTAACAGCTGACCTTATTTGGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 242
                QY 241 CTGATGTCGAATGTCGAAGCCCATTTAAATGTGAGCGAAATTCATGCAAGTCTTTC 300
                DB 243 CTGATGTCGAATGTCGAAGCCCATTTAAATGTGAGCGAAATTCATGCAAGTCTTTC 302
                QY 301 TGGCAGACGTGGGAAGATCTTTATCAACAGCAGATTTTCCAAACGACACAGTCAGCCCCGAG 360
                DB 303 TGGCAGACGTGGGAAGATCTTTATCAACAGCAGATTTTCCAAACGACACAGTCAGCCCCGAG 362
                QY 361 CCTCCGATCGTACTATGAAACAGTATCACTCACTAAATGAATCTATTCTTGGATGAAT 420
                DB 363 CCTCCGATCGTACTATGAAACAGTATCACTCACTAAATGAATCTATTCTTGGATGAAT 422
```


Db 1503 AGTCTAGTTACCTACCTTTCTTTGATTTTTCGACGTTTGACCTAGCCATCTCAAGCACT 1562
Qy 1398 TTGACGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAAAGATCATCTCAAGCTGATCAT 1457
Db 1563 TTGACGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAAAGATCATCTCAAGCTGATCAT 1622
Qy 1458 TGGATCTACTCAACAAAGGAAGGTTGTCAGAGTACATTAAGATTTCTGCTCAAA 1517
Db 1623 TGGATCTACTCAACAAAGGAAGGTTGTCAGAGTACATTAAGATTTCTGCTCAAA 1682
Qy 1518 TTTTCAATAAATTTCTTCTCTCTCTTTTAAATAAAAAAAAAAAAAA 1563
Db 1683 TTTTCAATAAATTTCTTCTCTCTTTTAAATAAAAAAAAAAAAAA 1728

RESULT 6
LOCUS I16100 1749 bp DNA linear PAT 03-APR-1996
DEFINITION Sequence 2 from patent US 5474901.
ACCESSION I16100
VERSION I16100.1 GI:1251008
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Antibodies to human carboxypeptidase B and methods of use thereof
JOURNAL Patent: US 5474901-A 2 12-DEC-1995;
FEATURES
source location/Qualifiers
1. 1749
/organism="unknown"
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN

Query Match 87.2%; Score 1372; DB 6; Length 1749;
Best Local Similarity 90.3%; Pred. No. 2.6e-306;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
Qy 1 AGAAATTTGCTTTGGGATCAAGCTTTGACAGCTTTGACGCTTTGACCCATTTGTTCTCT 60
Db 24 AGAAATTTGCTTTGGGATCAAGCTTTGACAGCTTTGACGCTTTGACCCATTTGTTCTCT 83
Qy 61 TCTGTAGCAGCAGTGTCTTCGCGTTCAGAGTCCAGAGTTCAGCTGCTCTTCTAGAA 120
Db 84 TCTGTAGCAGCAGTGTCTTCGCGTTCAGAGTCCAGAGTTCAGCTGCTCTTCTAGAA 143
Qy 121 CCTCTAGGCAAGTTCAGTTCACAGATCTTACTACAGATATGAGATTTGTTCTTGGC 180
Db 144 CCTCTAGGCAAGTTCAGTTCACAGATCTTACTACAGATATGAGATTTGTTCTTGGC 203
Qy 181 AGCCGGTAAACAGTGTACCTTTATTTGTAAGAAACAAAGTCCATTTTGTAAATGCAAT 240
Db 204 AGCCGGTAAACAGTGTACCTTTATTTGTAAGAAACAAAGTCCATTTTGTAAATGCAAT 263
Qy 241 CTGATGTGCAATGTGAAAGCCCATTTAAATGTAGCGGAAATTCATGCAAGTGTCTTGC 300
Db 264 CTGATGTGCAATGTGAAAGCCCATTTAAATGTAGCGGAAATTCATGCAAGTGTCTTGC 323
Qy 301 TGGCAGACGTGGAAGATCTTATTAACAGCAGATTTTCCACGACAGTCCAGCCCCGAG 360
Db 324 TGGCAGACGTGGAAGATCTTATTAACAGCAGATTTTCCACGACAGTCCAGCCCCGAG 383
Qy 361 CCTCCGATCGTACTAGACAGTATCACTCACTTAATGAATCTATTCTTGGATAGAAAT 420
Db 384 CCTCCGATCGTACTAGACAGTATCACTCACTTAATGAATCTATTCTTGGATAGAAAT 443
Qy 421 TTATACTAGAGGCACTCTGATATGCTTACAAAAATCCACATTTGGATCTCTCAATTTGAGA 480
Db 444 TTATACTAGAGGCACTCTGATATGCTTACAAAAATCCACATTTGGATCTCTCAATTTGAGA 503
Qy 481 AGTACCCACTCTATGTTTTAAAGTTTTCTGGAAGAAACAAACAGCCAAAAATGCCATAT 540

Db 504 AGTACCCACTCTATGTTTTAAAGTTTTCTGAAAAAGAAACAAACAGCCAAAAATGCCATAT 563
Qy 541 GGAATTGACTGTGGAATCCATGCCAGAGAATGGAATCTCTCTGCTTTCTGCTTGGTTCA 600
Db 564 GGAATTGACTGTGGAATCCATGCCAGAGAATGGAATCTCTCTGCTTTCTGCTTGGTTCA 623
Qy 601 TAGGCCAT----- 608
Db 624 TAGGCCATATACTCAATTTCTATGGGTAATAGGGCAATATACCAATCTCCTGAGGCTTG 683
Qy 609 -----A 609
Db 684 TGGATTTCTATGTTATGCGCGTGGTTAATGTGGAACGGTTATGACTACTCATGGAAGAAAGA 743
Qy 610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATCGGAACAATCATTTGCATCGGAACAGACC 669
Db 744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATCGGAACAATCATTTGCATCGGAACAGACC 803
Qy 670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGTGTCATCCAGTTCTCATGCT 729
Db 804 TGAATAGGAACCTTTGCTTCCAAACACTGGTGTGAGGAAGTGTCATCCAGTTCTCATGCT 863
Qy 730 CGGAAACCTACTGTGGACTTTTATCCTGAGTCAGAACGAGAGTGAAGGCGAGTGGCTAGTT 789
Db 864 CGGAAACCTACTGTGGACTTTTATCCTGAGTCAGAACGAGAGTGAAGGCGAGTGGCTAGTT 923
Qy 790 TCTTTGAGAAAGAAATATCAACCGATTTAAAGCATATCATGAGTCATTTCTATCTCCAGC 849
Db 924 TCTTTGAGAAAGAAATATCAACCGATTTAAAGCATATCATGAGTCATTTCTATCTCCAGC 983
Qy 850 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAACGTGCTC 909
Db 984 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAACGTGCTC 1043
Qy 910 TAGTAGCCAGTGAAGCAGTTCTGCTTATTTGACAAAACTAGTAAATAACAGGTATACAC 969
Db 1044 TAGTAGCCAGTGAAGCAGTTCTGCTTATTTGAGAAAACTAGTAAATAACAGGTATACAC 1103
Qy 970 ATGGCCATCGCTCAGAAACCTTATACAGTCTCTGAGGTGGGACGATTTGATCTATG 1029
Db 1104 ATGGCCATCGCTCAGAAACCTTATACAGTCTCTGAGGTGGGACGATTTGATCTATG 1163
Qy 1030 ATTTGGGCATCAATAATTCGT----- 1050
Db 1164 ATTTGGGCATCAATAATTCGTTTTACAATTTGAACCTTCAGATACGGGCACATACGGATCT 1223
Qy 1051 -----TTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1097
Db 1224 TGCTGCCGAGCGTTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1283
Qy 1098 TAGCTTTGGCATGTCTATTAGGAATGTTTAAATGCCCTGATTTTATCATTTCTGCTTCCGTAT 1157
Db 1284 TAGCTTTGGCATGTCTATTAGGAATGTTTAAATGCCCTGATTTTATCATTTCTGCTTCCGTAT 1343
Qy 1158 TTTAAATTTACTGATTTCCAGCAAGACCAAAATCATTTGATCAGATTTATTTTAAAGTTTATC 1217
Db 1344 TTTAAATTTACTGATTTCCAGCAAGACCAAAATCATTTGATCAGATTTATTTTAAAGTTTATC 1403
Qy 1218 CGTAGTTTGTATAAAGATTTTCTATTCTTGGTTCTGTCTGAGAGAACCTTAATAAGTGTCT 1277
Db 1404 CGTAGTTTGTATAAAGATTTTCTATTCTTGGTTCTGTCTGAGAGAACCTTAATAAGTGTCT 1463
Qy 1278 ACTTTGCCATTAAGGCAGACTAGGTTTCATGCTCTTTTAACTCCCTTTTAAAAAAATTTGTA 1337
Db 1464 ACTTTGCCATTAAGGCAGACTAGGTTTCATGCTCTTTTAACTCCCTTTTAAAAAAATTTGTA 1523
Qy 1338 AGTCTAGTTTACCTACTCTTTTCTTTGATTTTTCGACGTTTTCAGCTAGGCCATCTCAAGCAACT 1397
Db 1524 AGTCTAGTTTACCTACTCTTTTCTTTGATTTTTCGACGTTTTCAGCTAGGCCATCTCAAGCAACT 1583
Qy 1398 TTGAGCGTTTGTAGCTAGCCATCTCAAGCAAGTTTAAATCAAGATTCATCTCAAGCTGATCAT 1457

Db 1584 TTGACGTTTGTGACATGACATCTCAAGCAAGTTTAAATCAAAAGATCATCTCAGCGTGATCAT 1643
Qy 1458 TGGATCTCTACTCAACAAAAGGAGGGTGTGAGAGTACATTAAGAGATTTCTGCTCCAAA 1517
Db 1644 TGGATCTCTACTCAACAAAAGGAGGGTGTGAGAGTACATTAAGAGATTTCTGCTCCAAA 1703
Qy 1518 TTTTCAATAAATTTCTTCTCTCTCTTTAAAAAAMMMMMMM 1563
Db 1704 TTTTCAATAAATTTCTTCTCTCTCTTTAAAAAAMMMMMMM 1749

RESULT 7
I33526
LOCUS I33526 1749 bp DNA linear PAT 06-FEB-1997
DEFINITION Sequence 2 from patent US 5593674.
ACCESSION I33526
VERSION I33526.1 GI:1824317
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.R. and Eaton,D.L.
TITLE Plasma carboxypeptidase
JOURNAL Patent: US 5593674-A 2 14-JAN-1997;
FEATURES Location/Qualifiers
source 1..1749
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN

Query Match 87.2%; Score 1372; DB 6; Length 1749;
Best Local Similarity 90.3%; Pred. No. 2.6e-306;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

Qy 1 AGAAAAATGCTGTGGGATGAAGCTTTGCGAGCTTGCAGCTTGTGACCTTGTGACCTTGTCTCT 60
Db 24 AGAAAAATGCTGTGGGATGAAGCTTTGCGAGCTTGCAGCTTGTGACCTTGTGACCTTGTCTCT 83
Qy 61 TCTGTGAGCAGCATGCTTCTGCGTTCAGAGTGCGCCAGATTCAGCTGCTCTTCTAGAA 120
Db 84 TCTGTGAGCAGCATGCTTCTGCGTTCAGAGTGCGCCAGATTCAGCTGCTCTTCTAGAA 143
Qy 121 CCTCTAGCGCAAGTTCAGATTTCTACAGAACTTTACTACAAATATGAGATTTCTCTGCG 180
Db 144 CCTCTAGCGCAAGTTCAGATTTCTACAGAACTTTACTACAAATATGAGATTTCTCTGCG 203
Qy 181 AGCGGTAACAGCTGACCTTATTTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
Db 204 AGCGGTAACAGCTGACCTTATTTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263
Qy 241 CTGATGTGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 300
Db 264 CTGATGTGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 323
Qy 301 TGGCAGACGTGGAGATCTTTATCAACAGCAGATTTCCAAACGACACAGTCAGCCCGAG 360
Db 324 TGGCAGACGTGGAGATCTTTATTCACAGCAGATTTCCAAACGACACAGTCAGCCCGAG 383
Qy 361 CCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTTCTTGGATGAAT 420
Db 384 CCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTTCTTGGATGAAT 443
Qy 421 TTATACTGAGAGCATCTCTGATATGCTTACAAAATCCAGATTCCTCATTTGAGA 480
Db 444 TTATACTGAGAGCATCTCTGATATGCTTACAAAATCCAGATTCCTCATTTGAGA 503
Qy 481 AGTACCCACTCTATGTTTTTAAAGTTCTGGAAGAAACAAACAGCCAAAATGCCATAT 540
Db 504 AGTACCCACTCTATGTTTTTAAAGTTCTGGAAGAAACAAACAGCCAAAATGCCATAT 563
Qy 541 GGATTTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTCTCTGCTTGTGTTCA 600

Db 564 GGATTTGACTGTGGAATCCATGCGCAGAGAATGGATCTCTCTGCTTCTCTGCTTGTGTTCA 623
Qy 601 TAGGCCAT----- 608
Db 624 TAGGCCATATAACTCAATCTATGGGATTAATAGGCAATATACCAATCTCTCTGAGGCTTG 683
Qy 609 -----A 609
Db 684 TGGATTTTCTATGTTATGCGCGTGGTTAAATCTGCGCGTTATGACTACTCATGGAAGAA 743
Qy 610 ATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAAACAATCATTCGATCGGAACAGACC 669
Db 744 ATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAAACAATCATTCGATCGGAACAGACC 803
Qy 670 TGAATAGCAACTTTGTCTCCAAACACTGTTGAGGAAGTGATCCAGTTCCTCATGCT 729
Db 804 TGAATAGGAACCTTTGTCTCCAAACACTGTTGAGGAAGTGATCCAGTTCCTCATGCT 863
Qy 730 CGGAAACCTACTGTGAGACTTTATCTCTGAGTCAGAACAGAGTGAAGGCGAGTGGCTAGTT 789
Db 864 CGGAAACCTACTGTGAGACTTTATCTCTGAGTCAGAACAGAGTGAAGGCGAGTGGCTAGTT 923
Qy 790 TCTTGAGAGAAATATCAACAGATTAAGCATATACAGCATGCAATTCATCTCCAGC 849
Db 924 TCTTGAGAGAAATATCAACAGATTAAGCATATACAGCATGCAATTCATCTCCAGC 983
Qy 850 ATATAGTGTTCATATTCCTATACAGAGTAAGCAAGCAAGCAAGTGAAGCAAGTGTCTC 909
Db 984 ATATAGTGTTCATATTCCTATACAGAGTAAGCAAGCAAGCAAGTGAAGCAAGTGTCTC 1043
Qy 910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACAGGTATACAC 969
Db 1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATAGTAAAAATACAGGTATACAC 1103
Qy 970 ATGCGCATGCTCAGAAAACCTTATACCTAGCTCCTGAGGTGGGAGATTTGATCTATG 1029
Db 1104 ATGCGCATGCTCAGAAAACCTTATACCTAGCTCCTGAGGTGGGAGATTTGATCTATG 1163
Qy 1030 ATTTGGGCATCAATATTGCT----- 1050
Db 1164 ATTTGGGCATCAATATTGCTTTTCAATTGAATTCGAGATACGGGCAATACGGATTC 1223
Qy 1051 -----TTACATCAAAACCCAGCTTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1097
Db 1224 TGCTGCGGAGCGCTTACATCAAAACCCAGCTTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1283
Qy 1098 TAGCTTGGCATGTCATAGGAATGTTTAAATGCCCTGATTTTATCATTTCTGCTCCGTAT 1157
Db 1284 TAGCTTGGCATGTCATAGGAATGTTTAAATGCCCTGATTTTATCATTTCTGCTCCGTAT 1343
Qy 1158 TTTTAAATTTACTGATTCAGCAAGCAAAATCATTTGATCAGATTAATTTTAAAGTTTATC 1217
Db 1344 TTTTAAATTTACTGATTCAGCAAGCAAAATCATTTGATCAGATTAATTTTAAAGTTTATC 1403
Qy 1218 CGTAGTTTGTAAAAAGATTTTCTCTATTCTCTGTTCTGTGAGAGAACCTTAATAAGTCT 1277
Db 1404 CGTAGTTTGTAAAAAGATTTTCTCTATTCTCTGTTCTGTGAGAGAACCTTAATAAGTCT 1463
Qy 1278 ACTTTGCCATTAAGGCGAGTACGGGTTTCATGCTTTTAAACCTTTTAAAAAATGTAA 1337
Db 1464 ACTTTGCCATTAAGGCGAGTACGGGTTTCATGCTTTTAAACCTTTTAAAAAATGTAA 1523
Qy 1338 AGTCTAGTTTACCTATCTTTTCTTGTGATTTTGAAGTTTGTAGCTAGCCATCTCAGCAACT 1397
Db 1524 AGTCTAGTTTACCTATCTTTTCTTGTGATTTTGAAGTTTGTAGCTAGCCATCTCAGCAACT 1583
Qy 1398 TCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAAAGATCATCTCACGCTGATCAT 1457
Db 1584 TCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAAAGATCATCTCACGCTGATCAT 1643
Qy 1458 TGGATCTCTCAACAAAAGGAGGTGTGAGAGTACATTAAGATTTTCTGCTCCAAA 1517
Db 1644 TGGATCTCTCAACAAAAGGAGGTGTGAGAGTACATTAAGATTTTCTGCTCCAAA 1703


```
Qy 970 ATGGCCATGGCTCAGAAACCTTATACCTAGTCTCCTGGAGGTGGGAGCATTTGGATCTATG 1029
Db 1088 ATGGCCATGGCTCAGAAACCTTATACCTAGTCTCCTGGAGGTGGGAGCATTTGGATCTATG 1147
Qy 1030 ATTTGGGCATCAATATATCGT----- 1050
Db 1148 ATTTGGGCATCAATATATCGT----- 1207
Qy 1051 -----TTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTTAAAA 1097
Db 1208 TGCTGCGGACGGTTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTTAAAA 1267
Qy 1098 TAGCTTGGCATGTCAATAGGAATGTTAAATGCCCTGATTTTATCATTTCTGCTTCCGTAT 1157
Db 1268 TAGCTTGGCATGTCAATAGGAATGTTAAATGCCCTGATTTTATCATTTCTGCTTCCGTAT 1327
Qy 1158 TTTAAATTTACTGATTCAGCAAGACCAATCAATGTATATCAGATTAATTTTAAAGTTTATC 1217
Db 1328 TTTAAATTTACTGATTCAGCAAGACCAATCAATGTATATCAGATTAATTTTAAAGTTTATC 1387
Qy 1218 CGTAGTTTTCATAAAGATTTTCTATTCTCTTGGTCTGTGTCAGAGACCTAATAAGTGCT 1277
Db 1388 CGTAGTTTTCATAAAGATTTTCTATTCTCTTGGTCTGTGTCAGAGACCTAATAAGTGCT 1447
Qy 1278 ACTTTGCCATTAAAGCAGACTAGGGTTTCATGTCTTTTACCTTTTAAAAAAATTTGTA 1337
Db 1448 ACTTTGCCATTAAAGCAGACTAGGGTTTCATGTCTTTTACCTTTTAAAAAAATTTGTA 1506
Qy 1338 AAGCTAGTTACCTATCTTTTCTTTTGAATTTTTCGACGTTTGAAGCATCTCAAGCAACT 1397
Db 1507 AAGCTAGTTACCTATCTTTTCTTTTGAATTTTTCGACGTTTGAAGCATCTCAAGCAACT 1566
Qy 1398 TTGACGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAAGCTGATCAT 1457
Db 1567 TTGACGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAAGCTGATCAT 1626
Qy 1458 TGGATCTCTACTCAACAAAAGGAGGGTGTGAGAGTACATTAAGATTTTCTGCTCCAAA 1517
Db 1627 TGGATCTCTACTCAACAAAAGGAGGGTGTGAGAGTACATTAAGATTTTCTGCTCCATA 1686
Qy 1518 TTTTCAATAAA 1528
Db 1687 AAAAAAAAAA 1697

RESULT 9
AR086324
LOCUS
DEFINITION
Sequence 1 from patent US 5985562.
ACCESSION
AR086324
VERSION
AR086324.1 GI:10013090
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 1272)
AUTHORS
Morser, M. John. and Nagashima, M.
TITLE
Method of detecting thrombotic disease risk associated with plasma
carboxypeptidase B polymorphisms
JOURNAL
Patent: US 5985562-A 1 16-NOV-1999;
FEATURES
Location/Qualifiers
source
1..1272
/organism="unknown"
BASE COUNT
375 a 269 c 271 g 357 t
ORIGIN

Query Match 58.4%; Score 918; DB 6; Length 1272;
Best Local Similarity 86.8%; Pred. No. 2e-201;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

Qy 18 ATGAAGCTTTGACGCTTGAGTCCTTGTATCCCATCTTCTCTGTGTGAGCAGCATGTC 77
Db 1 ATGAAGCTTTGACGCTTGAGTCCTTGTATCCCATCTTCTCTGTGTGAGCAGCATGTC 60
```

```
Qy 78 TTCCGCTTCCAGAGTGGCCAAAGTTCTAGCTGTCTCTTCCCTAGAACCTCTTAGCAAGTTCAA 137
Db 61 TTCCGCTTCCAGAGTGGCCAAAGTTCTAGCTGTCTCTTCCCTAGAACCTCTTAGCAAGTTCAA 120
Qy 138 GTTCTACAGAAATCTTACTACAAATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 197
Db 121 GTTCTACAGAAATCTTACTACAAATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 180
Qy 198 CTTATTGTGAAGAAAAAAACAAAGTCCATTTTGTGTAATATGATCTGATGTGCAATGTG 257
Db 181 CTTATTGTGAAGAAAAAAACAAAGTCCATTTTGTGTAATATGATCTGATGTGCAATGTG 240
Qy 258 AAAGCCCATTTTAAATGTGAGCGGAATTCATGTCAGTGTCTTGTGTCAGACGTTGGAAGAT 317
Db 241 AAAGCCCATTTTAAATGTGAGCGGAATTCATGTCAGTGTCTTGTGTCAGACGTTGGAAGAT 300
Qy 318 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCAGGCCCGCCGAGCTCCGATCGTACTAT 377
Db 301 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCAGGCCCGCCGATCGTACTAT 360
Qy 378 GAACAGTATCATCTCAATTAATGAAATCTATTTTGGATAGAAATTTATTAATCTGAGAGCAT 437
Db 361 GAACAGTATCATCTCAATTAATGAAATCTATTTTGGATAGAAATTTATTAATCTGAGAGCAT 420
Qy 438 CCTGATATGCTTACAAAATCCACATTTGGATCTCTTCTGCTGGAAGTACCCACTCTATGTT 497
Db 421 CCTGATATGCTTACAAAATCCACATTTGGATCTCTTCTGCTGGAAGTACCCACTCTATGTT 480
Qy 498 TTAAGGCTTTCTGGAAGAAACAAACAGCAAAATGCCATATGGAATTTGATCTGGAATC 557
Db 481 TTAAGGCTTTCTGGAAGAAACAAACAGCAAAATGCCATATGGAATTTGATCTGGAATC 540
Qy 558 CATGCCAGAGATGGAATCTCTCTGCTTTCTGCTGTGTGTTCTATAGGCCAT----- 608
Db 541 CATGCCAGAGATGGAATCTCTCTGCTTTCTGCTGTGTGTTCTATAGGCCATATAAATCAA 600
Qy 609 ----- 608
Db 601 TTCTATGGGATTAATAGGGCAATATACCAATCTCTCTGAGGCTTGTGGATTTCTATGTTATG 660
Qy 609 -----AATCGAATGTGGAGAAAG 626
Db 661 CCGGTGTTAAATGTGACGCTTATGACTACTCATCGAAGAAAGAAATCGAATGTGAGAAAG 720
Qy 627 AACCGTTCTTTCTATCGGAACAAATCAATGTCGGAACAGACCTGAATAGCAACTTTTGTCT 686
Db 721 AACCGTTCTTTCTATCGGAACAAATCAATGTCGGAACAGACCTGAATAGCAACTTTTGTCT 780
Qy 687 TCCAAACACTGGTGTGAGGAAGGTGATCCAGTTCTCTCATGCTCGGAACCTACTCTGGA 746
Db 781 TCCAAACACTGGTGTGAGGAAGGTGATCCAGTTCTCTCATGCTCGGAACCTACTCTGGA 840
Qy 747 CTTTATCTGAGTCAGAACCCAGAAAGTGAAGGCAAGTGTGCTAGTTTCTTGAAGAAATATC 806
Db 841 CTTTATCTGAGTCAGAACCCAGAAAGTGAAGGCAAGTGTGCTAGTTTCTTGAAGAAATATC 900
Qy 807 AACCAAGTTAAAGCATATACATGATGATTTCAATCTCCAGCATATATAGTGTTCATAT 866
Db 901 AACCAAGTTAAAGCATATACATGATGATTTCAATCTCCAGCATATATAGTGTTCATAT 960
Qy 867 TCCTATACAGAGTAAAGCAAGAACCATGAGGAATCTGTCTCTAGTGTGAGCAAGCA 926
Db 961 TCCTATACAGAGTAAAGCAAGAACCATGAGGAATCTGTCTCTAGTGTGAGCAAGCA 1020
Qy 927 GTTCTGTCTATTGACAAAACTAGTAAAAATACAGGTATACATACATGAGCAATGCTCAGAA 986
Db 1021 GTTCTGTCTATTGAGAAAACTAGTAAAAATACAGGTATACATACATGAGCAATGCTCAGAA 1080
Qy 987 ACCTTATACCTAGCTCTCTGAGGTGGGAGCATTTGATCTTATGATTTGGGCATCAATAT 1046
Db 1081 ACCTTATACCTAGCTCTCTGAGGTGGGAGCATTTGATCTTATGATTTGGGCATCAATAT 1140
```

```
QY 1047 TCGT-----TTAC 1054
Db 1141 TCGTTACAATTGAATTCAGATACGGGACATACCGGATCTTGTGCGGAGCGGTAC 1200
QY 1055 ATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGTCTCTAAATAGCTTGGCATGTCTAT 1114
Db 1201 ATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGTCTCTAAATAGCTTGGCATGTCTAT 1260
QY 1115 AGGAATGTTTAA 1126
Db 1261 AGGAATGTTTAA 1272

RESULT 10
BD106653
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1272)
Morser, M.J. and Nagashima, M.
METHOD OF DETECTING THROMBOTIC DISEASE RISK
Patent: JP 2002502253-A 1 22-JAN-2002;
SCHERING AG
COMMENT
PN JP 2002502253-A/1
PD 22-JAN-2002
PF 02-JUN-1998 JP 1999501465
PR 03-JUN-1997 US 08/869057
PI MICHAEL JOHN MORSER, MARIKO NAGASHIMA
PC C12Q1/68
CC Strandedness: Double;
Toplogy: Linear;
FH Key Location/Qualifiers.
source
1. 1272
/organism="synthetic construct"
/mol.type="genomic DNA"
/db.xref="taxon:32630"
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN

Query Match 58.4%; Score 918; DB 6; Length 1272;
Best Local Similarity 86.8%; Pred. No. 2e-201;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

QY 18 ATGAAGCTTTGCAGCCTTGAGCTTGTACCAATGTTCTTCTGTGTGAGCAGCATGTC 77
Db 1 ATGAAGCTTTGCAGCCTTGAGCTTGTACCAATGTTCTTCTTCTGTGTGAGCAGCATGTC 60
QY 78 TTCGGCTTCAGAGTGCCCAAGTCTAGCTCTTCTAGAACCTCTAGGCAAGTTCAA 137
Db 61 TTCGGCTTCAGAGTGCCCAAGTCTAGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
QY 138 GTTCTACAGAACTTACTACAAATGAGATTGTTCTTGGCAGCGGTAACAGCTGAC 197
Db 121 GTTCTACAGAACTTACTACAAATGAGATTGTTCTTGGCAGCGGTAACAGCTGAC 180
QY 198 CTTATTGTGAAGAAAAACCAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAAATGTG 257
Db 181 CTTATTGTGAAGAAAAACCAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAAATGTG 240
QY 258 AAAGCCCAATTAATGTAGCGGAATTCATGAGTGTCTTGTGCGAGCGTGGAGAT 317
Db 241 AAAGCCCAATTAATGTAGCGGAATTCATGAGTGTCTTGTGCGAGCGTGGAGAT 300
QY 318 CTTATTCAACAGCAGATTTTCCACGACACAGTCAGCTCAGCCCGGAGCTCCGCATCGTACTAT 377
Db 301 CTTATTCAACAGCAGATTTTCCACGACACAGTCAGCTCAGCCCGGAGCTCCGCATCGTACTAT 360
```

```
QY 378 GAACAGTATCCTCAGTAAATGAATCTATTCTTGGATAGATTTTATACTAGAGGCAT 437
Db 361 GAACAGTATCCTCAGTAAATGAATCTATTCTTGGATAGATTTTATACTAGAGGCAT 420
QY 438 CTTGATATGCTTACAAAATCCACATTTGGATCTCTATTTCAGAAAGTACCCCACTCTATGTT 497
Db 421 CTTGATATGCTTACAAAATCCACATTTGGATCTCTATTTCAGAAAGTACCCCACTCTATGTT 480
QY 498 TTAAGGTTTCTGAAAAGAAACAAACAGCCAAAATATGCCATATGGATTGATCTGTGGAAATC 557
Db 481 TTAAGGTTTCTGAAAAGAAACAAACAGCCAAAATATGCCATATGGATTGATCTGTGGAAATC 540
QY 558 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTGTTTCATAGCCAT- 608
Db 541 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGCCATATAAATCAA 600
QY 609 ----- 608
Db 601 TTCTATGGGTAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 660
QY 609 -----AATCGAATGTGGAGAAAG 626
Db 661 CCGGTGTTAAATGTGGACGTTTATGACTACTCATGGAAGAAAGAAATCGAATGTGGAGAAAG 720
QY 627 AACCGTTCTTTCTATGCGAAACAATCATTTGCAATCGGAACAGACCTGAAATAGCAACTTTGTC 686
Db 721 AACCGTTCTTTCTATGCGAAACAATCATTTGCAATCGGAACAGACCTGAAATAGCAACTTTGTC 780
QY 687 TCCAAACACTGTTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGA 746
Db 781 TCCAAACACTGTTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGA 840
QY 747 CTTTATCTCAGTCAGAACCCAGAAAGTGAAGCGAGTGGCTAGTTTCTTGAAGAAATATC 806
Db 841 CTTTATCTCAGTCAGAACCCAGAAAGTGAAGCGAGTGGCTAGTTTCTTGAAGAAATATC 900
QY 807 AACGATTAAAGCATATCATGAGTGCATTTATCTCCAGCATATAGTGTGTTTCCATAT 866
Db 901 AACGATTAAAGCATATCATGAGTGCATTTATCTCCAGCATATAGTGTGTTTCCATAT 960
QY 867 TCCTATACAGGAATGAAGCAAGACCATGAGGAAGTGTCTCTAGTACGAGTGAAGCA 926
Db 961 TCCTATACAGGAATGAAGCAAGACCATGAGGAAGTGTCTCTAGTACGAGTGAAGCA 1020
QY 927 GTTCTGCTATTGACAAAACCTAGTAAAAATACCAAGGTATACACATGGCCATGGCTCAGAA 986
Db 1021 GTTCTGCTATTGAGAAAACCTAGTAAAAATACCAAGGTATACACATGGCCATGGCTCAGAA 1080
QY 987 ACCTTATACCTAGCTCTGAGAGTGGGAGCGATTGGATCTATGTTGGGCATCAAAATAT 1046
Db 1081 ACCTTATACCTAGCTCTGAGAGTGGGAGCGATTGGATCTATGTTGGGCATCAAAATAT 1140
QY 1047 TCGT-----TTAC 1054
Db 1141 TCGTTTACAATTGAACCTTCAGATACGGGCACATACCGAATTTTGTGCGGAGCGGTAC 1200
QY 1055 ATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGTCTCTAAATAGCTTGGCATGTCTAT 1114
Db 1201 ATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGTCTCTAAATAGCTTGGCATGTCTAT 1260
QY 1115 AGGAATGTTTAA 1126
Db 1261 AGGAATGTTTAA 1272

RESULT 11
BT006936
LOCUS
DEFINITION
Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA, complete cds.
ACCESSION
VERSION
BT006936.1 GI:30582710
KEYWORDS
FLI_CDNA.
```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Kalinine.N., Chen.X., Rolfs.A., Halleck.A., Hines.L., Eisenstein.S.,
Koundinya.M., Raphael.J., Moreira.D., Kelley.T., Labaer.J., Lin.Y.,
Phelan.M. and Farmer.A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1272)
AUTHORS Kalinine.N., Chen.X., Rolfs.A., Halleck.A., Hines.L., Eisenstein.S.,
Koundinya.M., Raphael.J., Moreira.D., Kelley.T., Labaer.J., Lin.Y.,
Phelan.M. and Farmer.A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: <http://bioinfo.clontech.com/orfclones>.
Location/Qualifiers
1..1272
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GH00156X1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
CDS
1..1272
/codon_start=1
/product="carboxypeptidase B2 (plasma, carboxypeptidase
u)"
/protein_id="AAP35582.1"
/db_xref="GI:30582711"
/translations="MKCSLAVLPVLFCEQHVFAFQSGQVLAALPRTSRQVQLQN
LTTTEIVLWQPTADLIKKQVHFVNASVDNKAHLNVSIGPCSVLLADVEDLI
QQQISNDTSPRASASYEQYHSLSNIYEWIEFITERHPDMLTKIHIGSFSEKYPFLY
LKVSQKQAKNAIWDICGTHAREWISPAFLWFIQHTQFYGIHQYTNLLRLVDY
VMPVNVVDGYDSWKNRMRNRSFVANNHCIGTDLNRFASKHWCBEASSSCSSE
TYCGLYPESEPEVKAVASFLRNINOIKAVISMHSYSQHIYFVPSYTRSKSDHELS
LVASEAVRAIEKTSKNTRYTHGHSETLILAPGGDDDIYDLGKYSFTIELRDTGT
GFLLPRIYKPTCREAPAAVASKIAHWIRNV"
BASE COUNT 373 a 268 c 273 g 358 t
ORIGIN
Query Match 58.2%; Score 915.4; DB 9; Length 1272;
Best Local Similarity 86.74; Pred. No. 8.1e-201;
Matches 1102; Conservative 0; Mismatches 6; Indels 163; Gaps 2;
QY 18 ATGAAGCTTTGCAGCTTGCAGTCCCTGTGACCAATGTTCTCTCTGTGTGAGCAGCATGTC 77
DB 1 ATGAGCTTTGCAGCTTGCAGTCCCTGTGACCAATGTTCTCTCTGTGTGAGCAGCATGTC 60
QY 78 TTCGCGTTCAGAGTGCCCAAGTCTAGTGTCTCTCTAGAACCTCTAGGCAAGTTCAA 137
DB 61 TTCGCGTTTCAGAGTGCCCAAGTCTAGTGTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
QY 138 GTTCTACAGATCTTACTACAAATATGAGATGTTCTCTGCGAGCGCGGTAAAGCTGAC 197
DB 121 GTTCTACAGATCTTACTACAAATATGAGATGTTCTCTGCGAGCGCGGTAAAGCTGAC 180
QY 198 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTGCAATGTG 257
|||||
|||||

181 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTGCAATGTG 240
258 AAAGCCCATTTAAATCTGAGCGGAATTCATGCAAGTCTCTGCTGGCAGACGTGGAGAT 317
241 AAAGCCCATTTAAATCTGAGCGGAATTCATGCAAGTCTCTGCTGGCAGACGTGGAGAT 300
318 CTTATTCAACAGCAGAGATTTCCAAACGACACAGTCAAGCCCCGAGCCCTCCGATCGTACTAT 377
301 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAAGCCCCGAGCCCTCCGATCGTACTAT 360
378 GAACAGTATCACTCACTAAATGAATCTATTCTTTGGATAGAAATTTATACTAGAGGCAT 437
361 GAACAGTATCACTCACTAAATGAATCTATTCTTTGGATAGAAATTTATACTAGAGGCAT 420
438 CTTGATATGCTTACAAAATTCACATTTGGATCTCTATTGGAGAGTACCCACTCTATGTT 497
421 CTTGATATGCTTACAAAATTCACATTTGGATCTCTATTGGAGAGTACCCACTCTATGTT 480
498 TTAAGGTTTCTGAAAAAGAAACAAACAGACCAAAAATGCCATATGGATTGACTGTGGAATC 557
481 TTAAGGTTTCTGAAAAAGAAACAAACAGACCAAAAATGCCATATGGATTGACTGTGGAATC 540
558 CATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCAT----- 608
541 CATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATATAAATCAA 600
609 ----- 608
601 TTTATATGGGTAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 660
609 -----AATCGAATGTGGAGAAAG 626
661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGAAAAAAGATCGAATGTGGAGAAAG 720
627 AACCGTTTCTTTATCGAAACAATCATTCGATCGGAACAGACCTGAATAGCAACTTTGTC 686
721 AACCGTTTCTTTATCGGAACAATCATTCGATCGGAACAGACCTGAATAGCAACTTTGCT 780
687 TCCAAACTGTTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
781 TCCAAACTGTTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840
747 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCAAGTGGCTAGTTTCTTTGAGAAGAAATATC 806
841 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCAAGTGGCTAGTTTCTTTGAGAAGAAATATC 900
807 AACCAATTAAAGCATATCATCATGATGCATTCATATCCAGCATATAGTGTGTTTCCATAT 866
901 AACCAATTAAAGCATATCATCATGATGCATTCATATCCAGCATATAGTGTGTTTCCATAT 960
867 TCCTATACAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 926
961 TCCTATACAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1020
927 GTTCTGCTTATTGACAAAATCTAGTAAAAATACCAAGGTATACATACATGCCCCATGGCTCAGAA 986
1021 GTTCTGCTTATTGAGAAAATCTAGTAAAAATACCAAGGTATACATACATGCCCCATGGCTCAGAA 1080
987 ACCTTATACCTAGCTCCTGAGGTGGGACGATTTGGATCTATGATTTGGGCAATCAATAT 1046
1081 ACCTTATACCTAGCTCCTGAGGTGGGACGATTTGGATCTATGATTTGGGCAATCAATAT 1140
1047 TCCT-----TTAC 1054
1141 TCCTTATCAATTTGAACCTTCAGATACGGGCACATACGGATTTCTTGTGCCGAGCGGTAC 1200
1055 ATCAAAACCCACCTGTGAGAGGCTTTTTCGCGTGTCTCTAAAAATAGTCTGGCATGTCTATT 1114
1201 ATCAAAACCCACCTGTGAGAGGCTTTTTCGCGTGTCTCTAAAAATAGTCTGGCATGTCTATT 1260
1115 AGGAATGTTTA 1125
1261 AGGAATGTTTA 1271
DB

```

RESULT 12
BT007990
LOCUS
DEFINITION
    1272 bp mRNA linear SYN 13-MAY-2003
    Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
    carboxypeptidase U) mRNA, partial cds.
ACCESSION
BT007990
VERSION
BT007990.1 GI:30584818
KEYWORDS
FLI CDNA.
SOURCE
    synthetic construct
    ORGANISM
        Homo sapiens
        carboxypeptidase B2 (plasma,
        carboxypeptidase U)
REFERENCE
    1 (bases 1 to 1272)
    Kalline, N., Chen, X., Rolfs, A., Halleck, A., Hines, L., Eisenstein, S.,
    Koundinya, M., Raphael, J., Moreira, D., Kelley, T., LaBaer, J., Lin, Y.,
    Phelan, M., and Farmer, A.
    Cloning of human full-length CDSs in BD Creator(TM) System Donor
    vector
    JOURNAL
    Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
    Circle, Palo Alto, CA 94303, USA
    This CDS clone is a part of a collection of human full length
    expression clones generated by BD Biosciences Clontech and the
    Harvard Institute of Proteomics. Each CDS has been cloned in two
    forms: with and without stop-codon (to allow fusion with C-terminal
    tag). The CDS has been directionally cloned using BD In-Fusion(TM)
    cloning system between the SalI and HindIII sites of the pDNR-DUAL
    vector. Additional sequences in the clone: 'ACC' after SalI site
    and before 'ATG' to provide Kozak consensus sequence; 'GG' after
    last codon and before HindIII site to maintain reading frame.
    Clone distribution: http://bioinfo.clontech.com/orfclones.
    Location/Qualifiers
        1..1272
            /organism="synthetic construct"
            /mol_type="mRNA"
            /db_xref="taxon:32630"
            /clone_lib="BD Creator(TM) CDS Library derived from MGC
            collection"
            /lab_host="DH5alpha T1 resistant"
            /notes="Vector: pDNR-Dual"
            1..>1272
                /notes="Mutations: 1271:Stop->Leu"
                /codon_start=1
                /transl_table=11
                /product="Homo sapiens carboxypeptidase B2 (plasma,
                carboxypeptidase U)"
                /protein_id="AA036662.1"
                /db_xref="GI:30584819"
                /translation="MKLCSLAVLPVILFCEQHFVAFQSGQVLAALPRTSRQVQLQN
                LITTVIRVLQPTADLIKKQVHFVNASDVNDVKHNLNSGIPCSVLLADVEDLI
                QOOISNDTVSPRASVYQYHSLNIYSWIEPIITERHPDMLTKIHIGSSPEKYPLVY
                LKVSQEAQKNAWIDCGIHAREWISPAFLWFGHITQFYGLIQYINLLVDVY
                VMPVNVNDYDYSKQNRWNRKNRSFYANNHCIGTDLNRFASKHWCBSGSSSCSE
                TYGLYPESEPEVKAVSLRRNINQIKAYISMHSYQHVFPYSYTRSKSDHELS
                LVASEAVRAIEKTSKTRTHYGHGSETLYLAPGGGDDWIYDLGIKISYFIELRDITGY
                GFLLPRTYKPTCREAFVAKIAWIRNVL"
                372 a 268 c 273 g 359 t
BASE COUNT
    372 a 268 c 273 g 359 t
ORIGIN
    78 TTGCGGTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTTAGGCAAGTCAA 137
    61 TTGCGGTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTTAGGCAAGTCAA 120
    138 GTTCTACAGAACTTACTACACATATGAGATGTTCTCTGGCAGCGGGTAACAGCTGAC 197
    121 GTTCTACAGAACTTACTACACATATGAGATGTTCTCTGGCAGCGGGTAACAGCTGAC 180
    198 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTTGTAAATGCACTCTGATGTCGACATGTG 257
    181 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTTGTAAATGCACTCTGATGTCGACATGTG 240
    258 AAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGTCGACAGCTGGAAGAT 317
    241 AAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGTCGACAGCTGGAAGAT 300
    318 CTTATTTCACAGCAGATTTTCCACGACACAGTCCAGCCCCCGAGCCTCCGATCGTACTAT 377
    301 CTTATTTCACAGCAGATTTTCCACGACACAGTCCAGCCCCCGAGCCTCCGATCGTACTAT 360
    378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 437
    361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 420
    438 CTTGATATCTTACAAAAATCCACATTTGGATCCTCATTTTGAAGATACCCACTCTATGTT 497
    421 CTTGATATCTTACAAAAATCCACATTTGGATCCTCATTTTGAAGATACCCACTCTATGTT 480
    498 TTAAAGGTTTCTGGAAGAAACAACAGACGCAAAAATGCCATATGGATTGATCTGGAATC 557
    481 TTAAAGGTTTCTGGAAGAAACAACAGACGCAAAAATGCCATATGGATTGATCTGGAATC 540
    558 CATGCCAGAGATGGATCTCTCTGCTTCTTCTGCTGTTTCTATGAGCCATATTAATCA 608
    541 CATGCCAGAGATGGATCTCTCTGCTTCTTCTGCTGTTTCTATGAGCCATATTAATCA 600
    609 ----- 608
    601 TTCTATGGGTAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTCTATGTTATG 660
    609 -----AATCGAATGTGGAGAAAG 626
    661 CCGTGTTTAAATGTGGATGTTATGACTACTCATGGAAGAAAGATCGAATGTGGAGAAAG 720
    627 AACCGTTCTTCTATCGGAACAATCATTCGATCGGAACAGACCTGAATAGCAACTTTGTC 686
    721 AACCGTTCTTCTATCGGAACAATCATTCGATCGGAACAGACCTGAATAGCAACTTTGTC 780
    687 TCCAAACACTGTTGTGAGGAAGTGCAATCCAGTTCTCTGCTCGGAACCTTACTGTGGA 746
    781 TCCAAACACTGTTGTGAGGAAGTGCAATCCAGTTCTCTGCTCGGAACCTTACTGTGGA 840
    747 CTTTATCTCTGAGTCAGAACACAGAAAGTGAAGGAGTGGCTAGTTCTTGTGAGAAATATC 806
    841 CTTTATCTCTGAGTCAGAACACAGAAAGTGAAGGAGTGGCTAGTTCTTGTGAGAAATATC 900
    807 AACCGATTAAAGCATACATCAGATGCAATCTATCTCCAGCATATAGTGTGTTTCCATAT 866
    901 AACCGATTAAAGCATACATCAGATGCAATCTATCTCCAGCATATAGTGTGTTTCCATAT 960
    867 TCCATATACGAGAGTAAAGCAAGAACCATGAGGAACTGTCTAGTAGGCAAGTGAAGCA 926
    961 TCCATATACGAGAGTAAAGCAAGAACCATGAGGAACTGTCTAGTAGGCAAGTGAAGCA 1020
    927 GTTCTGCTATTGACAAAATCTAGTAAAAATACAGATATACATGGCCATGCTCAGAA 986
    1021 GTTCTGCTATTGACAAAATCTAGTAAAAATACAGGATATACATGGCCATGCTCAGAA 1080
    987 ACCTTATACCTAGCTCCTCGAGTGGGAGCATTTGATCTTATGTTGGGCATCAATAT 1046
    1081 ACCTTATACCTAGCTCCTCGAGTGGGAGCATTTGATCTTATGTTGGGCATCAATAT 1140

```

```

Query Match      58.1%; Score 914.4; DB 12; Length 1272;
Best Local Similarity 86.7%; Pred. No. 1.4e-200;
Matches 1101; Conservative 0; Mismatches 6; Indels 163; Gaps 2;

QY 18 ATGAAGCTTTGACGCTTGCAGTCCCTTGATCCCATGTTCTCTCTGTTGAGGAGCATGTC 77
DB 1 ATGAAGCTTTGACGCTTGCAGTCCCTTGATCCCATGTTCTCTCTGTTGAGGAGCATGTC 60

```

```
QY 1047 TCCT-----TTAC 1054
|||
Db 1141 TCCTTACATTGAACCTTCAGATACGGGCACATACGGATTCCTTGCTGCCGGAGGTTAC 1200
|||
QY 1055 ATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTCTAAATAGCTTGGCATGTCAAT 1114
|||
Db 1201 ATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTCTAAATAGCTTGGCATGTCAAT 1260
|||
QY 1115 AGGAATGTTT 1124
|||
Db 1261 AGGAATGTTT 1270
|||

RESULT 13
LOCUS AB021968 Mus musculus mRNA for carboxypeptidase R, complete cds. ROD 27-JUL-2000
DEFINITION AB021968 1421 bp mRNA
ACCESSION AB021968
VERSION AB021968.1 GI:9558447
KEYWORDS carboxypeptidase R.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Obata,K., Okada,N.,
Campbell,W. and Okada,H.
TITLE Pro-carboxypeptidase R is an acute phase protein in the mouse,
whereas carboxypeptidase N is not
JOURNAL J. Immunol. 165 (2), 1053-1058 (2000)
MEDLINE 20341711
PUBMED 10878383
REFERENCE 2 (bases 1 to 1421)
AUTHORS Sato,T.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1998) Tomoo Sato, Fukushima Hospital, Chofu
Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi
441-8124, Japan (E-mail:tomoo@tcp-ip.or.jp, Tel:81-532-46-7511,
Fax:81-532-46-4899)
FEATURES
source Location/Qualifiers
1. .1421
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
17..1285
/codon_start=1
/product="carboxypeptidase R"
/protein_id="BAB03402.1"
/db_xref="GI:9558448"
/translation="MKLHGIILVAILLYEHLGFQSFQSVLALPRTSRQVOLLQNL
TTTYEVLWQPVTAETIEKKEVHFVNASDVSVKAHLNVSRIPIFNVLNNVEDLIE
QQFNDTVSPRASAYEYOYHSLEIYSWTEVIEQHPDMLQIKIYIGSSFEKPYLV
KVGSKQRIKNALWIDCGIHAREWISPAFLWFIQVYTFPHKENLYTRLLRHVDYFI
MPVNVVDYDYTWKQNRMRKNRSKHNKRCVGTDLNRFNPSKHKWCKEKGSSSCSE
YGLYPESEBEVKAVADFLRNIDHIKAYISMESYSOQILFPYSYNRSKSKDHEELSL
VASENRAIESINKNRYTHGSSSESLYLAPEGSDDWIDYDLGIKYSFTELRTGRY
FLPERYIKETCAEALAAISKIVHWIRNT"
polyA_site 1421
/notes="11 a nucleotides"
BASE COUNT 418 a 320 c 305 g 378 t
ORIGIN
Query Match 41.7%; Score 656; DB 10; Length 1421;
Best Local Similarity 72.2%; Pred. No. 6.7e-141;
Matches 1009; Conservative 0; Mismatches 215; Indels 174; Gaps 5;
QY 4 AAATGCTGTGGATGAAGCTTTGCAGCTTGCGAGTCCTTGATCCCATGTTCTCTCT 63
|||
Db 3 AAGTACTGTGGATGAAGCTTCATGCGCTTGAATCCCTGGTAGCCATCATCTCTTA-- 60
|||
QY 64 GTGAGCAGCATGCTTCGCGCTTCCAGAGTGCCCAAGTTCTAGCTGCTCTTCCTAGAACCT 123
|||
Db 61 -TGAGCAGCATGCTTCGCGCTTCCAGAGTGCCCAAGTTTATCTGCTCTTCCAGAACCT 119
|||
```

```
124 CTAGGCAAGTTCAAGTTCTACAGAAATCTTACTACAAATATGAGATGTTCTCTCGCAGC 183
|||
120 CCAGGCAGGTTCAACTACTTCAAGATCTTACTACAAAGTATGAGGTCGTTCTCTGGCAGC 179
|||
184 CGGTAAACAGCTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTGTAAATGCAATCTG 243
|||
180 CAGTGACAGCTGAATTCATCGAGAGAAAAAGGAAGTCCCACTTTTGTGAAATGCGCTCG 239
|||
244 ATGTCGACAAATGTGAAAGCCATTTAAATGTGAGCGGAATTCATCGAGTGTCTTGTGTG 303
|||
240 ATGTCGACAGTGTCAAAGCGCATTTAAATGTGAGCAGAAATTCATTTAAAGTCTGATGA 299
|||
304 CAGACGTGGAAGATCTTATTCAACGACAGATTTCCAAACGACACAGTCAGCCCCGAGCCT 363
|||
300 ACAACGTGGAGGACCTTAATTTGAACGACAGACTTTCAATGACCGGTGAGCCCCGCGCCT 359
|||
364 CGGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTA 423
|||
360 CGGCTTCACTATGAGCAGTATCACTCGTAAATGAATCTATTCTCGATAGAAATCA 419
|||
424 TAACTGAGAGGCAATCTGATATGCTTTACAAAATCCACATTTGGATCCTCATTTGAGAAGT 483
|||
420 TAACTGAACAGCATCTCGATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGT 479
|||
484 ACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAACAACAGCCAAAATGCCATATGA 543
|||
480 ACCCACTTATGTTTAAAGGTTCTAGGAAAGGAACAAGAATCAAAAATGCCATCTGA 539
|||
544 TTGACTGTGGAATCCATGCGCAGAGATGATCTCTCTGTTCTGTTGTGTTGTTGTTGTTG 603
|||
540 TCGACTGTGGAATCCATGCGCAGAGATGATTTCACTGCTTCTGTTGTGTTGTTGTTGTTG 599
|||
604 GC----- 605
|||
600 GCTACGTGACACAATTCATGGGAAAGAAAATCTGTATACAGACTTCTGAGGACAGTGG 659
|||
606 -----CATATC 612
|||
660 ATTTCTACATCATGCCCGTGTGAACGTGTGATGACTATGACTACAGTGGAAAAAGAAATC 719
|||
613 GAATGTGAGAGAAAGAACCGTTCTTTCTATGCGAAACAATCATTTGCATCGGAACAGACTGA 672
|||
720 GAATGTGAGAGAAAGAACCGTTCTGCTCAACAAGAACACCGCTCGGTGGGACAGACTGA 779
|||
673 ATAGCAACTTTGTCTCAAACACTGTTGTGAGGAAGGTGATCCAGTTCTCTATGCTCGG 732
|||
780 ACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCTCTGCTCTG 839
|||
733 AAACCTTACTGTGACCTTTATCTCTGACTCAGAACCAAGTGAAGGCGAGTGGCTAGTTTCT 792
|||
840 AAACCTTACTGTGAGCTTTATCTCTGAGTGTGAGCGAGGTTGAAGGAGTGGCTAGTTTCT 899
|||
793 TGAGAGAAATATCAACACAGATTTAAAGCATATACATGACATGCTATCTATCTCCAGCAT 852
|||
900 TGAGAGAAATATCGACCAATTTAAAGCTTACATGATGACTATCTATCTATCTCCACAAA 959
|||
853 TAGTGTTCATATTCCTATFACAGGAAGTAAAGCAAGAACCAATGAGGAACCTGCTCTCTAG 912
|||
960 TACTGTTTCCCTATTCTATTAACAGAAAGCAAGAACCAAGGACCAAGGAAACTGCTCTCTAG 1019
|||
913 TAGCCAGTGAAGCAGTTGCTGCTATTGACAAAATCTAGTAAAAATACCAGGTATACACATG 972
|||
1020 TGCCAGCAGGAAGCAGTTGCTGCAATTTGAAAGTATTAATAAAACACCAAGGTACACACG 1079
|||
973 GCCATCGCTCAGAAAACCTTATACCTAGCTCCTCGAGGTGGGAGCAGTTGGATCTATGATT 1032
|||
1080 GCAGTGGCTCAGAAAAGTTTATATCTAGCTCCTGAGGTTCTGACGATGATCTATGATT 1139
|||
1033 TGGGCATCAAAATATTCGTT----- 1051
|||
1140 TGGGCATCAAAATATTCGTTTACAATTGAGTCCGAGATACAGGCAGATACGGATTCCTTGC 1199
|||
```



```

QY 1052 -----TACATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTCTAATATAG 1100
Db 1200 TGCTGAGAGATACATCAAAACCCACCTTGTGAGAGCTTTGGCGCATCTCTAATATAG 1259
QY 1101 CTTGGCATGTCATTAGGAATGTTTAAATGCCCTGATTTTATCATCTGCTTCGGTATTTT 1160
Db 1260 TTTGGCATGTCATCAGGAACACITTAATGCC-----TAACCTCCGCTCTCATATTTT 1313
QY 1161 AATTACTGATTCAGCAAGCAACAAATCATGTATCAGAT--TATTTTAAAGTTTATCC 1218
Db 1314 ATTTTATTGATTTTCAGCAACACITTAACCTGTGATAGCTTCTAAGTTGAATCAGTTCC 1373
QY 1219 GTAGTTTGTGATAAAGAT 1236
Db 1374 TTGTTTTTGTGAAGAAT 1391

RESULT 14
AF164524
LOCUS AF164524 1490 bp mRNA linear ROD 05-APR-2000
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)
mRNA, complete cds.
ACCESSION AF164524
VERSION AF164524.1 GI:7416966
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1490)
Marx,P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van
Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
Characterization of mouse thrombin-activatable fibrinolysis
inhibitor
Thromb. Haemost. 83 (2), 297-303 (2000)
MEDLINE 20201996
PUBMED 10739389
REFERENCE 2 (bases 1 to 1490)
Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
Direct Submission
Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100,
Utrecht 3584 CX, The Netherlands
JOURNAL Location/Qualifiers
source
1. 1490
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:477035"
1. 1490
/gene="Tafi"
84. 1352
/function="carboxypeptidase"
/codon_start=1
/product="thrombin-activatable fibrinolysis inhibitor"
/protein_id="AAF62385.1"
/db_xref="GI:7416967"
/translations="MKLHGILVAILVEHQHGFAGQVLSALPRTSRVOLLONL
TTTVEVLQPVTAETLEKKEVHFVFNASDVSVKHLNVSIPFNVLNNVEDLIE
QTFNFTVSPRASATYEQHSLNEIYVIEBQPDMLQKIYIGSSFEKPLPLVL
KVGSKQRINKAIWIDCGIHAREWISPAFCLVFTQPHGKENVYLRLLRHVDYI
PMVMYDGYDTWKRMWRKNRSASHNNRCVGTDLNRNFAKHWCKEKGASSSCSET
YCGLYPSEPEVAVADFLRNIDHIKAVISMHSYSQQLIFPYSYNRSKSKDHEELSL
VASEAVRAIESINKNRYTHGSGESLSLAPGSDWDWIDLGIKYSFTIELRDTGRY
FLLPERYKPTCREALAAISKIVHWIRNT"
BASE COUNT 438 a 337 c 322 g 393 t
ORIGIN
Query Match 41.7%; Score 656; DB 10; Length 1490;
Best Local Similarity 72.2%; Pred. No. 6,6e-141;
Matches 1009; Conservative 0; Mismatches 215; Indels 174; Gaps 5;
QY 4 AATATGCTGTGGATGAAGCTTTGCAGCCTTGACGCTTCCTTGACCAATGTTCTCTTCT 63

```

```

Db 70 AAGTCACTGTGTGGATGAAGCTTCATGGCCCTTGAATCTGGTAGGCATCATCTCTA-- 127
QY 64 GTGAGCAGCATGCTTTCGGGTTCCAGAGTGGCCAAAGTTCTAGTGTCTTCTCTAGAACCT 123
Db 128 -TGAGCAGCATGCTTTCGGGTTCCAGAGTGGCCAAAGTTTATCTGCTCTTCCAGAACCT 186
QY 124 CTAGGCAAGTTCAAGTCTTACAGAACTCTTACTACAAACATATGAGATTTGTTCTCTGGCAGC 183
Db 187 CCAGGCAAGTTCAACTACTTACAGAACTTCTTACTACAAAGTATGAGGCTGTTCTCTGGCAGC 246
QY 184 CGGTACACGCTGACCTTATTTGTGAAGAAACAAAGTCCATTTTGTGTAATGCATCTG 243
Db 247 CAGTGACAGCTGAATTCATCGAAGAAAGAAAGTCCACTTTTGTGAATGCGCTG 306
QY 244 ATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAGTGTCTTGTGG 303
Db 307 ATGTCGACAGTGTCAAAAGCGCATTTAATGTGAGCAGAAATTCATTTAAGTTCATGA 366
QY 304 CAGACGTGGAAGATCTTTATTCACACAGAGATTTCCAAACGACACAGTCAGCCCCGAGCCT 363
Db 367 ACAACGTGGAGGACCTTAATTTGAACAGCAGACTTTTCAATGACACGGTCAGCCCCGCGCT 426
QY 364 CCGCATCGTACTATGAACAGATATCACTCAATGAATGAATCTATTCTTGTGATGAATTA 423
Db 427 CCGCTTCATATATGAGCAGTATCACTCGCTAAATGAATCTATTCTTGTGATGAAGTCA 486
QY 424 TAACTCAGAGGACCTCTGATATGCTTTACAAAATTCACATTTGATGATCTCTATTTGAGAAGT 483
Db 487 TAACTGACAGCATCTCTGACATGCTCCAGAAATCTACATCGGATCATCATTCGAGAGT 546
QY 484 ACCACTCTATGTTTAAAGGTTTCTGAAAGAAACAAACAGACCAAAATGCCATATGGA 543
Db 547 ACCACTTTATGTTTAAAGGTTCTCAGGAAAGGAAACAAAGAAATCAAAATGCCATCTGGA 606
QY 544 TTGACTGTGGAATCCATGCGCAGAGATGATCTCTCTGCTTCTTGTGTTGTTCTATAG 603
Db 607 TCAGCTGTGGAATCCATGCGCAGAGATGGAATTTTACCTGCTTTCTGTTGTTGTTCTATAG 666
QY 604 GC----- 605
Db 667 GCTACGTGACACAAATTCATGGAAGAAATCTGTATACCAGACTTCTGAGGCACGTGG 726
QY 606 -----CATATC 612
Db 727 ATTTCTACATCATGCCCGGTGATGAACGTGATGCTATGACTACACGTGGAAGAAATC 786
QY 613 GAATGTGAGAGAAACCGTTCTTTCTATGCGGAACAAATCATTCATCGCGGAACAGACTGA 672
Db 787 GAATGTGAGAGAAACCGCTCTGCTTCAAAAGAAACAAACCGCTCGTGGGCACAGACCTGA 846
QY 673 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGG 732
Db 847 ACAGGAACCTTCGTTCCAAACACTGGTGTGAGGAAGTGGTCAAGTTCCTCTGCTCTG 906
QY 733 AAACCTACTGTGGACTTTATCTGAGTCAGAACCAAGTGAAGGAGGAGTGGCTAGTTTCT 792
Db 907 AAACCTACTGTGGACTTTATCTGAGTCTGAGCCAGAGGTGAAGGAGTGGCTGACTTCT 966
QY 793 TGAGAAAGAAATATCAACACAGATTAAGCATATCATATGAGTATGATCTACTCCAGATA 852
Db 967 TGAGAAAGAAATATCGACCAACATTAAGCTTTACATCATGATGCACTCATACTCCCAACA 1026
QY 853 TAGTGTTCCTATTTCTTATACACGAAGTAAAGCAAGCAACCATGAGGAGTCTCTCTAG 912
Db 1027 TACTGTTTCCCTTATTCCTATAACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1086
QY 913 TAGCCAGTGAAGCAGTTCGTGCTTATTTGACAAACTAGTAAATAACAGGTATACACATG 972
Db 1087 TGCCAGCGAAGCAGTTCGTGCAATTTGAAGTATTAATAAACAACAGGTATACACACG 1146
QY 973 GCATCGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGACGAGTATGATCTATGATT 1032

```

```
Db 1147 GCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAGATT 1206
Qy 1033 TGGGCATCAATATTCGTT----- 1051
Db 1207 TGGGCATCAATATTCGTTTACAAATTGAGCTCCGAGATACAGGCAGATACGGATTCCTGC 1266
Qy 1052 -----TACATCAAAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAATAG 1100
Db 1267 TGCTGAGAGATACATCAACCCACTTGTGCAGAGCTTTGGCGCCATCTCTAAATAG 1326
Qy 1101 CTGGCAGTCATPAGGAATGTTTAATGCCCTGATTTTATCAATCTGCTTCCGATTTT 1160
Db 1327 TTTGGCATGTATCAGGACACATTAATGCC-----TAACCTCCGCTCTCAATATTTT 1380
Qy 1161 AATTACTGATTCAGCAGACCAATCATGTATCAGAT--TATTTTAAAGTTTATCC 1218
Db 1381 ATTTTATGATTCAGCAACACTTAATCTGTGCAATAGCTTCTAAAGTTGAATCAGTTTC 1440
Qy 1219 GTAGTTTGTATAAAGAT 1236
Db 1441 TTGGTTTGGTGAAGAT 1458

RESULT 15
AF186188
LOCUS Mus musculus carboxypeptidase U mRNA, complete cds.
DEFINITION
ACCESSION AF186188
VERSION AF186188.1 GI:6003651
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1430)
He,Y.C. and Broze,G.
Isolation and characterization of mouse liver carboxypeptidase B
gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1430)
He,Y.C. and Broze,G.
Direct Submission
Submitted (13-SEP-1999) Division of Hematology, Washington
University School of Medicine, 216 S. Kingshighway, St. Louis, MO
63110, USA
FEATURES
Location/Qualifiers
1..1430
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6/CBAF1J"
/db_xref="taxon:10090"
20..1288
/notes="pro-carboxypeptidase-B; thrombin-activatable
fibrinolysis inhibitor (TAPI); pCPU; pCPB; binds
plasminogen"
/codon_start=1
/product="carboxypeptidase U"
/db_xref="GI:6003652"
/translational="MKLHGLGILVAILLYEQHGFQFQGVLSALPRTSRQVOLLQNL
TTTVEVLVQPVTAETFEKKKHFVFNASDVSVKALNVSRIPFNLMNVNVEDLIE
QQTNDVSPRASAYVEQVHSLNEIYSWIEVLTEQHPDMLQIKIYGSSEPKYPLVL
KVGSEQRINKYDIWKNNRWRKNSAHKNNRCVGDLDNRNFSKHWCKEKGASSSSCSE
MPVNVVDGYITWKNNRWRKNSAHKNNRCVGDLDNRNFSKHWCKEKGASSSSCSE
YGLYPESEPEKAVADFLRNTHIKAYISMSYSQOILFPYSYNRSKSKDHEELSL
VASEAVRAIGSFNKNTRYTHGSGSESLYAPGGSDWIYDLGIKYSPTIELRDTGRYG
FLLPERYIKPTCEALAAISKIVHVRNT"
BASE COUNT 421 a 320 c 310 g 379 t
ORIGIN

Query Match : 41.6%; Score 654; DB 10; Length 1430;
Best Local Similarity 72.4%; Pred. No. 1.9e-140;
Matches 989; Conservative 0; Mismatches 205; Indels 172; Gaps 4;
```

```
Qy 1 AGAAAAATGCTGTTGGATGAAGCTTTTGAGCGCTTGCAGTCCCTTGTACCCATTTGTTCTCT 60
Db 3 AGAAAGTCACTGTTGGATGAAGCTTTCATGGCCCTTGGAAATCCCTGGTAGCCATCATCTCT 62
Qy 61 TCTGTGAGCAGCATGTCTTTCGGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTTCTCTAGAA 120
Db 63 A---TCAGCAGCATGCTTTCGGCTTTCAGAGTGGCCAGGTTTATCTGCTCTTCCAAGAA 119
Qy 121 CTTCTAGGCAAGTTCAAGTCTTACAGAACTTACTACAACATATGAGATTGTTCTCTGGC 180
Db 120 CTTCCAGGCAAGTTCAACTTCTAGAACTTCTACTACAAGTATGAGTGGTCTCTCTGGC 179
Qy 181 AGCCGGTAAACAGCTGACCTTTATTTGCAAGAAAAACAAGTCCCAATTTTGTGTAATGCAT 240
Db 180 AGCAGTGACAGCTGAATTCATCGAGAGAAAAAGAGTCCCACTTTTGTGAAATGCGT 239
Qy 241 CTGATGTCGAATGTGAAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 300
Db 240 CTGATGTCGACAGTGTCAAAGCGCAATTTAAATGTGAGCAGAAATCCATTTAACGTTCTGA 299
Qy 301 TGGCAGACGTGGAAGATCTTTTCAACAGCAGAGATTTTCAACGACACAGCAGTCCGCCGCG 360
Db 300 TGAACCAACGTGGAGGACCTTAATTTGAACAGCAGACTTTTCAATGACCGGTCCAGCCCGCG 359
Qy 361 CCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTTCTTGGATAGAAT 420
Db 360 CTTCCGCTTCATACTATGAGAGTATCACTCGGTAAATGAAATCTATTTCTTGGATAGAG 419
Qy 421 TTATAACTGAGAGGCACTCTGATATGCTTACAAAAATCCACATTTGGATCTCTATTTGAGA 480
Db 420 TCTTAACCTGAACAGCATCTTGACATGCTCCAGAAAAATCTACATCGGATCATCATTCGAGA 479
Qy 481 AGTACCCACTCTATGTTTAAAGTCTTGGAAAAAGAAACAAACAGCCAAAAATGCCATAT 540
Db 480 AGTACCCACTCTTATGTTTAAAGGCTCTCAGGAAAGGAACAAAGAAATCAAAAAATGCCATCT 539
Qy 541 GGATTCACCTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTTCTGCTGCTGGTTCA 600
Db 540 GGATCCACTGTGGAATCCATGCCAGAGATGGAAATTTACCTGCTTTCTGTTGTTGTTCA 599
Qy 601 TAGGC----- 605
Db 600 TAGGCTACGTGACACAAATTCATGGGAAGAAAAATCTGTATACCAGACTTCTGAGGCAAG 659
Qy 606 -----CAT 609
Db 660 TGGATTTTACATCATGCTCCCGTGATGAACGTTGGATGGCTATGACTACACGTGAAAAAGA 719
Qy 610 ATCGAATGTGGAGAAAAAGAACCGTTCTTCTATCGGAACAATCATTTGCATCGGAACAGACC 669
Db 720 ATCGAATGTGGAGAAAGAACCGCTCTGCTCAAGAAGAACCCCTGGTGGGGCAGACACC 779
Qy 670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATGCT 729
Db 780 TGAACAGGAATTCGCTTCCAAACACTGGTGTGAGAAAGTGGTCAAGTTCTCTCTGCT 839
Qy 730 CGAAACCTACTCTGAGCTTTATCTCTGAGTCAGAACGAGAGTGAAGGAGTGGCTAGTT 789
Db 840 CTGAAACCTACTCTGAGCTTTATCTCTGAGTCTGAGCCAGAGGTGAAGGCAAGTGGCTGACT 899
Qy 790 TCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGCTATCTATCTATCTCCAGC 849
Db 900 TCTTGAGAGAAATATCGACCACTTAAGCTTACATCAGTATGCACTCATCTATCTCCAAC 959
Qy 850 ATATAGTGTTCCTATATTCCTATACAGAGTAAAAAGCAAGACCATGAGGAATCTGTCTC 909
Db 960 AATAGTGTTCCTATTTCTATAACAGAGCAAAAGCAGGACCCAGGAAGAACTGTCTC 1019
Qy 910 TAGTAGCCAGTGAAGCAGTTCTGCTATTGACAAAACTAGTAAAAATACAGGTATACAC 969
Db 1020 TAGTGGCCAGCAAGCAGTCTGTCGTAATTGGAAAGTTTATAAAAAACACCAGGTACACAC 1079
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:40:19 ; Search time 461.88 Seconds
(without alignments)
9193.336 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattgctgtggatg.....aaaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 510512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1573	100.0	1573	21	Human brain carbox
2	1469.8	93.4	1625	25	Human protease CDN
3	1372	87.2	1728	24	Gene #2397 used to
4	1372	87.2	1749	17	Human plasma carbo
5	1372	87.2	1749	18	Human plasma carbo
6	1370.4	87.1	1749	14	Human plasma carbo
7	918	58.4	1272	20	Human plasma carbo
8	853.8	54.3	1400	21	Lung cancer associ

9	478.6	30.4	55827	25	ABX13671	Human protease gen
10	331.2	21.1	416	25	ABX49430	Bovine EST associa
11	324	20.6	333	20	AAH85957	Human single nucle
12	324	20.6	333	20	AAH87390	Human single nucle
13	286.8	18.2	397	25	ABX43260	Bovine EST associa
14	241.6	15.4	251	19	AAH11374	Human biallelic po
15	167	10.6	231	24	ABN19133	Human ORFX polymu
16	137.2	8.7	431	24	ABK62561	Rat sequence diffe
17	124	7.9	927	17	AAAT35760	Rat mature carboxy
18	123.4	7.8	921	16	AAQ90601	Porcine carboxypep
19	123.4	7.8	1215	16	AAQ90600	Porcine Tyr-His-Me
20	120	7.6	121	20	AAH85710	Human single nucle
21	120	7.6	121	20	AAH85711	Human single nucle
22	109	6.9	1263	19	AAZ17195	Human pancreatic c
23	107.6	6.8	2023	25	ABV75116	Anti-human seminal
24	107.4	6.8	999	17	AAAT42500	Mature HCPB coding
25	107.4	6.8	1053	17	AAAT42497	Mature HCPB-(His)6
26	107.4	6.8	1263	17	AAAT42494	Human pancreatic c
27	107.4	6.8	1284	17	AAAT42506	ProHCPB gene with
28	104.2	6.6	1053	20	AAAT424804	Human carboxypepta
29	104.2	6.6	1059	17	AAAT42511	Modified HCPB (D25
30	104.2	6.6	1059	17	AAAT42512	Modified HCPB (D25
31	104.2	6.6	1059	18	AAAT62787	Carboxypeptidase B
32	104.2	6.6	1059	18	AAAT62788	Carboxypeptidase B
33	102.6	6.5	1059	18	AAAT62805	Carboxypeptidase B
34	102.6	6.5	1059	18	AAAT62806	Carboxypeptidase B
35	102.6	6.5	1059	18	AAAT62791	Carboxypeptidase B
36	102.6	6.5	1059	18	AAAT62801	Carboxypeptidase B
37	101	6.4	1059	18	AAAT62789	Carboxypeptidase B
38	101	6.4	1059	18	AAAT62790	Carboxypeptidase B
39	101	6.4	1870	18	AAAT62792	Human encoding huma
40	101	6.4	1870	19	AAV41819	Human carboxypepti
41	101	6.4	2154	18	AAV17331	PreproHCPB-linker-
42	99.4	6.3	1059	18	AAAT62808	Carboxypeptidase B
43	99.4	6.3	1059	18	AAAT62804	Carboxypeptidase B
44	99.2	6.3	1059	18	AAAT62800	Carboxypeptidase B
45	97.8	6.2	1059	18	AAAT62807	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAC81962
ID AAC81962 standard; cDNA; 1573 BP.

XX AC AAC81962;

XX DT 01-MAR-2001 (first entry)

XX DE Human brain carboxypeptidase B cDNA.

XX KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX CDS 18..1100
XX FT /*tag= a
XX FT /product= "carboxypeptidase B"

XX PN WO200066717-A1.

XX PD 09-NOV-2000.

XX PF 01-MAY-2000; 2000WO-JP02878.

XX PR 30-APR-1999; 99JP-0125169.

XX

(MATS/) MATSUMOTO A.
Matsumoto A;
WPI; 2000-687534/67.
P-PSDB; AAB11457.
Human brain carboxypeptidase B isolated from the hippocampus useful for screening agents for the treatment of Alzheimer's and other brain disorders -
Claim 2b; Page 64-68; 84pp; Japanese.
This invention describes a novel protein with peptidase activity against brain beta-amyloid precursor protein which has been isolated from human hippocampus and which has cerebroprotective, antialzheimer's, CC nootropic, neuroprotective and hemostatic activity and which can be used CC as a vaccine or for gene therapy. The protein, and compounds identified CC by screening as promoters or inhibitors of its activity, are used to CC regulate beta-amyloid accumulation in the brain and treat or prevent CC diseases in which this occurs, such as Alzheimer's, senile dementia, CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;
Query Match 100.0%; Score 1573; DB 21; Length 1573;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGAAATTTGCTTGGGATGAAGCTTTGAGCCCTTGAGCTCTTGTACCCATTGTTCTCT 60
Db 1 AGAAATTTGCTTGGGATGAAGCTTTGAGCCCTTGAGCTCTTGTACCCATTGTTCTCT 60
Qy 61 TCTGTAGAGCAGTCTTCCGCTTCCAGATGCGCCCAAGTTCAGCTGCTCTTCTAGAA 120
Db 61 TCTGTAGAGCAGTCTTCCGCTTCCAGATGCGCCCAAGTTCAGCTGCTCTTCTAGAA 120
Qy 121 CCTCTAGGCAAGTCAAGTCTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGC 180
Db 121 CCTCTAGGCAAGTCAAGTCTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGC 180
Qy 181 AGCGGTAAACAGCTGACCTTTATTTGGAAGAAAAACAAAGTCCATTTTTTTGTAATGCAT 240
Db 181 AGCGGTAAACAGCTGACCTTTATTTGGAAGAAAAACAAAGTCCATTTTTTTGTAATGCAT 240
Qy 241 CTGATGTGCAATGTGAAAGCCCAATTTAAATGTGAGCGGAATTCATGCAAGTCTTTCG 300
Db 241 CTGATGTGCAATGTGAAAGCCCAATTTAAATGTGAGCGGAATTTCCATGCAAGTCTTTCG 300
Qy 301 TGGCAGACGTGGAGATCTTATTCAACAGCAGATTTCCACGACACAGTCAGCCCCGAG 360
Db 301 TGGCAGACGTGGAGATCTTATTCCACAGCAGATTTCCACGACACAGTCAGCCCCGAG 360
Qy 361 CCTCCGATCGTACTATCAACAGTATCACTCACTAAATGAAATCTATTTTGGATAGAT 420
Db 361 CCTCCGATCGTACTATCAACAGTATCACTCACTAAATGAAATCTATTTTGGATAGAT 420
Qy 421 TTATACTGAGAGCAGTCTGATATGCTTACAAATTCACATTCATTCATTTGAGA 480
Db 421 TTATACTGAGAGCAGTCTGATATGCTTACAAATTCACATTCATTCATTTGAGA 480
Qy 481 AGTACCCACTGATGTTTAAAGTTTCTGGAAGAACCAACAGCCAAAATGCCATAT 540
Db 481 AGTACCCACTGATGTTTAAAGTTTCTGGAAGAACCAACAGCCAAAATGCCATAT 540
Qy 541 GGATTGACTGTGGAATCCATGCCAGAGATGATCTCTCTGCTTTTCTGCTTGTGTTCA 600
Db 541 GGATTGACTGTGGAATCCATGCCAGAGATGATCTCTCTGCTTTTCTGCTTGTGTTCA 600
Qy 601 TAGGCCAATATCGAATGTGAGAAAGAACCGTCTTCTTATGCGAAACAATCATTTGCATCG 660
Db 601 TAGGCCAATATCGAATGTGAGAAAGAACCGTCTTCTTATGCGAAACAATCATTTGCATCG 660

Qy 661 GAACAGACCTGAATAGCAACTTTGTCTCCAAACACCTGGTGTGAGGAAGGTGCATCCAGTT 720
Db 661 GAACAGACCTGAATAGCAACTTTGTCTCCAAACACCTGGTGTGAGGAAGGTGCATCCAGTT 720
Qy 721 CCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCCAGAAAGTGAAGCAG 780
Db 721 CCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCCAGAAAGTGAAGCAG 780
Qy 781 TGGCTAGTTTCTTGAGAAGAAATATCAACAGATTAAAGCATACATCAGCATGCTCATTCAT 840
Db 781 TGGCTAGTTTCTTGAGAAGAAATATCAACAGATTAAAGCATACATCAGCATGCTCATTCAT 840
Qy 841 ACTCCAGCATATAGTGTTCATATTTCTATACAGAAAGTAAAGCAAGACCATGAGG 900
Db 841 ACTCCAGCATATAGTGTTCATATTTCTATACAGAAAGTAAAGCAAGACCATGAGG 900
Qy 901 AACTGTCTCTAGTAGGCAAGTGTGCTATTGACAAAACCTAGTAAATAACCA 960
Db 901 AACTGTCTCTAGTAGGCAAGTGTGCTATTGACAAAACCTAGTAAATAACCA 960
Qy 961 GGTATACATGCGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGAGATT 1020
Db 961 GGTATACATGCGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGAGATT 1020
Qy 1021 GGATCTATGATTTGGGCATCAAAATATTCGTTTACATCAAAACCCACCTGTAGAGAAGCTTT 1080
Db 1021 GGATCTATGATTTGGGCATCAAAATATTCGTTTACATCAAAACCCACCTGTAGAGAAGCTTT 1080
Qy 1081 TGCCGCTGTCTTAAATAGCTTGGCATGTCAATTAGGAATGTTTAAATGCCCTGATTTTA 1140
Db 1081 TGCCGCTGTCTTAAATAGCTTGGCATGTCAATTAGGAATGTTTAAATGCCCTGATTTTA 1140
Qy 1141 TCATCTGCTCGTATTTTAAATTTACTGATCCAGCAAGCAAAATCATTTGATCAGAT 1200
Db 1141 TCATCTGCTCGTATTTTAAATTTACTGATCCAGCAAGCAAAATCATTTGATCAGAT 1200
Qy 1201 TATTTTAAAGTTTATCCGTTAGTTTGTATAAAGATTTTCTTATCTTGGTTCTGTGAG 1260
Db 1201 TATTTTAAAGTTTATCCGTTAGTTTGTATAAAGATTTTCTTATCTTGGTTCTGTGAG 1260
Qy 1261 AGAACCTAATAAGTGTCTACTTTGCCATTAAAGCAGACTAGGGTTCAITGCTTTTACCCT 1320
Db 1261 AGAACCTAATAAGTGTCTACTTTGCCATTAAAGCAGACTAGGGTTCAITGCTTTTACCCT 1320
Qy 1321 TTAATAAAAAATGTAAGTCTAGTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
Db 1321 TTAATAAAAAATGTAAGTCTAGTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
Qy 1381 AGCCATCTCAAGCAACTTTTCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGAT 1440
Db 1381 AGCCATCTCAAGCAACTTTTCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGAT 1440
Qy 1441 CATCTCACGCTGATCATTTGGATCCTACTCAACAAAAAGGAGGTGTGAGAGTACATTA 1500
Db 1441 CATCTCACGCTGATCATTTGGATCCTACTCAACAAAAAGGAGGTGTGAGAGTACATTA 1500
Qy 1501 AAGATTTCTGCTCCAAATTTCAATAAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1560
Db 1501 AAGATTTCTGCTCCAAATTTCAATAAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1560
Qy 1561 AAAAAAATAAAAA 1573
Db 1561 AAAAAAATAAAAA 1573
RESULT 2
ID ABX13670
XX ABX13670 standard; cDNA; 1625 BP.
AC ABX13670;
XX
DT 14-FEB-2003 (first entry)
XX

DE Human protease cDNA.
XX Human; gene; ss; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
XX single nucleotide polymorphism; SNP.
OS Homo sapiens.
XX Key Location/Qualifiers
FH 1..16
FT 5'UTR /*tag= a
FT CDS 17..1177
FT /*tag= b
FT /*product= "Protease"
FT variation replace(521,A)
FT /*tag= c
FT /*standard_name= "Single nucleotide polymorphism"
FT 3'UTR 1178..1625
FT /*tag= d
XX US2002137179-A1.
PN 26-SEP-2002.
XX 21-MAR-2001; 2001US-0813133.
XX 21-MAR-2001; 2001US-0813133.
XX (GANW/) GAN W.
PA (KETC/) KETCHUM K A.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX Gan W, Ketchum KA, Di Francesco V, Beasley EM;
XX WPI; 2003-102387/09.
DR P-PSDB; ABG72093, ABG72095.
XX New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
PT -
XX Claim 4; Fig 1; 80pp; English.
XX The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for
CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorders. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC cDNA, the gene for which is located on chromosome 13.
XX Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;

Query Match 93.4%; Score 1469.8; DB 25; Length 1625;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1553; Conservative 0; Mismatches 17; Indels 53; Gaps 2;
QY 2 GAAAATTGCTGTGGGATGAAGCTTTGCAGCCTTGCAGTCTTGTACCCATTTCTCTCTT 61
DB 1 GAAAATTGCTGTGGGATGAAGCTTTGCAGCCTTGCAGTCTTGTACCCATTTGTCTCTT 60
QY 62 CTGTGAGCAGCATGTCTTTCGCGTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTCTCTAGAAC 121
DB 61 CTGTGAGCAGCATGTCTTTCGCGTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTCTAGAAC 120
QY 122 CTCTAGGCAAGTTCAAGTTCTACAGAACTTTACTACAAATATGAGATTTCTCTGGCA 181
DB 121 CTCTAGGCAAGTTCAAGTTCTACAGAACTTTACTACAAATATGAGATTTCTCTGGCA 180
QY 182 GCCGTTAAACAGCTGACCTTTATGTGAAGAAAACAAAGTCCATTTTGTAAATGCATC 241
DB 181 GCCGTTAAACAGCTGACCTTTATGTGAAGAAAACAAAGTCCATTTTGTAAATGCATC 240
QY 242 TGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAATGTCTTGCT 301
DB 241 TGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAATGTCTTGCT 300
QY 302 GGCAGACGTGGAAGATCTTTATTCACAGCAGATTTCCACGACACAGTTCAGCCCCGAGC 361
DB 301 GGCAGATGTGAAGATCTTTATTCACAGCAGATTTCCACGACACAGTTCAGCCCCGAGC 360
QY 362 CTCGGATGCTACTATGAACAGTATCACTCACTAAATGAATCTATCTTGGATAGAAAT 421
DB 361 CTCGGATGCTACTATGAACAGTATCACTCACTAAATGAATCTATCTTGGATAGAAAT 420
QY 422 TATAACTGAGAGGATCTCTGATATGCTTACAAAATCCATTTGGATCTCTCATTTGAGAA 481
DB 421 TATAACTGAGAGGATCTCTGATATGCTTACAAAATCCATTTGGATCTCTCATTTGAGAA 480
QY 482 GTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAAACAGCCCAAAATGCCATATG 541
DB 481 GTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAAACAGCCCAAAATGCCATATG 540
QY 542 GATTGATGTTGGAATCCATGCGCAGAGAAATGATCTCTCTGCTTTCTGCTGTTGTTGAT 601
DB 541 GATTGATGTTGGAATCCATGCGCAGAGAAATGATCTCTCTGCTTTCTGCTGTTGTTGAT 600
QY 602 AGCCCATATCGAATCTGGAGAAAGACCGTTCTTCTATGCGAAACATCATTTGCATCGG 661
DB 601 AGCCCATATCGAATCTGGAGAAAGACCGTTCTTCTATGCGAAACATCATTTGCATCGG 660
QY 662 AACAGACCTGAATAGCAACTTTTCTCCAAACACTGTTGTGAGGAAGGTGCATCCAGTTC 721
DB 661 AACAGACCTGAATAGCAACTTTTCTCCAAACACTGTTGTGAGGAAGGTGCATCCAGTTC 720
QY 722 CTCATGCTCGAAACCTTACTGTGGAATTTATCTCTGAGTCAGAACACAGAAATGAAGCAGT 781
DB 721 CTCATGCTCGAAACCTTACTGTGGAATTTATCTCTGAGTCAGAACACAGAAATGAAGCAGT 780
QY 782 GGCATGTTCTTGAAGAAATATCAACAGATTAAGCATATCATGAGCATGATTCATA 841
DB 781 GGCATGTTCTTGAAGAAATATCAACAGATTAAGCATATCATGAGCATGATTCATA 840
QY 842 CTCACGACATATAGTGTTCATATTTCTTATACCAAGTAAAGCAAGCAACCATGAGGA 901
DB 841 CTCACGACATATAGTGTTCATATTTCTTATACCAAGTAAAGCAAGCAACCATGAGGA 900
QY 902 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGTGACAAATCTAGTAAAAATACCAG 961
DB 901 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGTAGAAAATTTAGTAAAAATACCAG 960
QY 962 GTATACATGCGCATGGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCAGATTG 1021
DB 961 GTATACATGCGCATGGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCAGATTG 1020


```
Db 603 TAGGCCATATAACTCAATCTTATGGGATATAGGCAATATACCAATCTCTGAGGCTTG 662
Qy 609 -----A 609
Db 663 TGGATTTCTATGTATATGCCGGTGGTTAAATGTGACGGTTATGACTACTCATGTGAAAAAGA 722
Qy 610 ATCGAATGTGAGAAAGACCGTTCTTCTATCGAACATCATTCGATCGGAACAGACC 669
Db 723 ATCGAATGTGAGAAAGACCGTTCTTCTATCGAACATCATTCGATCGGAACAGACC 782
Qy 670 TGAATAGCAACTTTGTCTCAAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
Db 783 TGAATAGCAACTTTGTCTCAAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 842
Qy 730 CGAAACCTACTGTGACATTTATCTCGAGTCAGAAACAGAAAGTGAAGCGAGTGGTAGTT 789
Db 843 CGAAACCTACTGTGACATTTATCTCGAGTCAGAAACAGAAAGTGAAGCGAGTGGTAGTT 902
Qy 790 TCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCATCTCCAGC 849
Db 903 TCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCATCTCCAGC 962
Qy 850 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGAACTGTCTC 909
Db 963 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGAACTGTCTC 1022
Qy 910 TAGTACGAGTGAAGCAGTTCGTGTATGACAAACTAGTAAATAACAGATATACAC 969
Db 1023 TAGTACGAGTGAAGCAGTTCGTGTATGACAAACTAGTAAATAACAGATATACAC 1082
Qy 970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGAGGATTCGATCTATG 1029
Db 1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGAGGATTCGATCTATG 1142
Qy 1030 ATTTGGGCATCAATATTCCTG----- 1050
Db 1143 ATTTGGGCATCAATATTCCTGTTACAAATGAACTTCGAGATACGGGCACATACGATTCCT 1202
Qy 1051 -----TTATCAACAAACCCACTGTAGAGAGCTTTTGGCGTGTCTCTAAAA 1097
Db 1203 TGCTGCGGAGCGTTATACATCAACCCACCTGTAGAGAGCTTTTGGCGTGTCTCTAAAA 1262
Qy 1098 TAGCTTGGCATGTGCATAGAAATGTTTAAATGCCCTGATTTTATCATTCGCTTCGGTAT 1157
Db 1263 TAGCTTGGCATGTGCATAGAAATGTTTAAATGCCCTGATTTTATCATTCGCTTCGGTAT 1322
Qy 1158 TTTAAATTTACTGATTCAGCAACCAAAATCATTTGTATCAGATTTATTTTAAAGTTTATC 1217
Db 1323 TTTAAATTTACTGATTCAGCAACCAAAATCATTTGTATCAGATTTATTTTAAAGTTTATC 1382
Qy 1218 CGTAGTTTGTATAAAGATTTTCTATTCCTTGGTTCTGTGAGAACCTTAATAAGTGCT 1277
Db 1383 CGTAGTTTGTATAAAGATTTTCTATTCCTTGGTTCTGTGAGAACCTTAATAAGTGCT 1442
Qy 1278 ACTTGGCATTAAGGAGACTAGGTTTCATGCTTTTACCTTTTAAAAAAATTTGTA 1337
Db 1443 ACTTGGCATTAAGGAGACTAGGTTTCATGCTTTTACCTTTTAAAAAAATTTGTA 1502
Qy 1338 AAGTCTAGTTTACCTACTTTTCTTCTGATTTTTCGACGTTTGTAGCCATCTCAAGCACT 1397
Db 1503 AAGTCTAGTTTACCTACTTTTCTTCTGATTTTTCGACGTTTGTAGCCATCTCAAGCACT 1562
Qy 1398 TTCGACGTTTGTAGTCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAAGCTGATCAT 1457
Db 1563 TTCGACGTTTGTAGTCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAAGCTGATCAT 1622
Qy 1458 TGGATCCTACTCAAAAAGGAGGGTGTGAGAGTACATTAAGATTTTGTCTCCAA 1517
Db 1623 TGGATCCTACTCAAAAAGGAGGGTGTGAGAGTACATTAAGATTTTGTCTCCAA 1682
Qy 1518 TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAATTTAAAAA 1563
Db 1683 TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAATTTAAAAA 1728
```

RESULT 4

```
AAT11671
ID AAT11671 standard; DNA; 1749 BP.
XX
AC AAT11671;
XX
XX 25-MAR-2003 (updated)
DT 12-APR-1996 (first entry)
XX
XX Human plasma carboxypeptidase B coding sequence.
XX
XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
CDS 41..1312
FT /*tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /*tag= b
FT mat_peptide 107..1309
FT /*tag= c
XX
XX US5474901-A.
XX
XX 12-DEC-1995.
XX
XX 19-JUL-1994; 94US-0277540.
XX
XX 01-FEB-1991; 91US-0649591.
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1996-039508/04.
DR P-PSDB; AAR90293.
XX
XX Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A
XX
XX Disclosure; Figure 4; 40pp; English.
XX
XX An antibody which specifically binds human plasma carboxypeptidase B
CC (hPCPB) and does not cross react with other carboxypeptidases is
CC useful for the detection of hPCPB in vitro. The antibody is also
CC used for purifying hPCPB from a sample. Purification comprises
CC passing a sample thought to contain hPCPB over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hPCPB.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
XX
XX Query Match 87.2%; Score 1372; DB 17; Length 1749;
XX Best Local Similarity 90.3%; Pred. No. 0;
XX Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
XX
Qy 1 AGAAATTCCTGTGGGATGAAGCTTTGCAGGCTTCAGTCTTGACCAATTTGTTCTCT 60
Db 24 AGAAATTCCTGTGGGATGAAGCTTTGCAGGCTTCAGTCTTGACCAATTTGTTCTCT 83
Qy 61 TCTGTAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAAGTTCTAGTCTCTTCCTAGAA 120
|||
```

```
Db 84 TCTGTGAGCAGCATGCTTCGCGCTTCAGAGTGGCCCAAGTTCTAGCTGTCTTCCTAGAA 143
QY 121 CCTCTAGGCAAGTTCAGATCTTACAGAACTCTTACTCAACATATGAGATTGTTCTCTGGC 180
Db 144 CCTCTAGGCAAGTTCAGATCTTACAGATCTTACTCAACATATGAGATTGTTCTCTGGC 203
QY 181 AGCCGGTAAACAGCTGACCTTTATTTGTGAAGAAAAACAAGTCCATTTTTTTGTAAATGCAAT 240
Db 204 AGCCGGTAAACAGCTGACCTTTATTTGTGAAGAAAAACAAGTCCATTTTTTTGTAAATGCAAT 263
QY 241 CTGATGTGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 300
Db 264 CTGATGTGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 323
QY 301 TGGCAGACGTGGAGATCTTTATTTCAACAGCAGATTTCCACGACACAGATCAGCCCGGAG 360
Db 324 TGGCAGACGTGGAGATCTTTATTTCAACAGCAGATTTCCACGACACAGATCAGCCCGGAG 383
QY 361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
Db 384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443
QY 421 TTATACTGAGAGGCATCTCGATATGCTTACAAAATTTCAAAAATTTCAAAAATTTCAAAAATTTCA 480
Db 444 TTATACTGAGAGGCATCTCGATATGCTTACAAAATTTCAAAAATTTCAAAAATTTCAAAAATTTCA 503
QY 481 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAAACAGCCAAATGCCATAT 540
Db 504 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAAACAGCCAAATGCCATAT 563
QY 541 GGATTGACTGTGGAATCCATGCGCAGAGATGGATCTCTCTGCTTCTGCTGTGGTTCA 600
Db 564 GGATTGACTGTGGAATCCATGCGCAGAGATGGATCTCTCTGCTTCTGCTGTGGTTCA 623
QY 601 TAGGCCAT----- 608
Db 624 TAGGCCATATACTCAATCTATGGGATAATAGGGCAATATACCAATCTCCTCAGGCTTG 683
QY 609 -----A 609
Db 684 TGGATTCTATGTTATGCGCGTGGTTAAATGTGACGGTTATGACTACTCATGAAAAAGA 743
QY 610 ATCGAATGTGAGAAAGAACCGTTCTTTCTATCGAACAAATCATTTGATCGGAACAGACC 669
Db 744 ATCGAATGTGAGAAAGAACCGTTCTTTCTATCGAACAAATCATTTGATCGGAACAGACC 803
QY 670 TGAATAGCAACTTTGTCTCAAAACACTGTGTGAGGAAGTGCATCCAGTTCCTCATGCT 729
Db 804 TGAATAGCAACTTTGTCTCAAAACACTGTGTGAGGAAGTGCATCCAGTTCCTCATGCT 863
QY 730 CGGAAACCTACTGTGGAATTTATCTGAGTCAGAACCCAGAAAGTGAAGGCGAGTGTAGTT 789
Db 864 CGGAAACCTACTGTGGAATTTATCTGAGTCAGAACCCAGAAAGTGAAGGCGAGTGTAGTT 923
QY 790 TCTTTGAGAGAAATATCAACAGATTAAGCATACATACATGATGCTATCTATCTCCAGC 849
Db 924 TCTTTGAGAGAAATATCAACAGATTAAGCATACATACATGATGCTATCTATCTCCAGC 983
QY 850 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAACTGTCTC 909
Db 984 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAACTGTCTC 1043
QY 910 TAGTAGCAGTGAAGAGTTCGTGCTATTGACAAAATAGTAAATAACAGGTATACAC 969
Db 1044 TAGTAGCAGTGAAGAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103
QY 970 ATGGCCATGCTCAGAAACCTTATACCTAGCTCCTGAGGTGGGACGATTGGATCTATG 1029
Db 1104 ATGGCCATGCTCAGAAACCTTATACCTAGCTCCTGAGGTGGGACGATTGGATCTATG 1163
QY 1030 ATTTGGGCATCAATATTCGT----- 1050
Db 1164 ATTTGGGCATCAATATTCGT----- 1223
```

```
QY 1051 -----TTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1097
Db 1224 TGTGTCGGGAGCGTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1283
QY 1098 TAGCTTGGCATGTCAATAGGAATGTTTAATGCCCCTGATTTTATCATCTTCTGCTCCGAT 1157
Db 1284 TAGCTTGGCATGTCAATAGGAATGTTTAATGCCCCTGATTTTATCATCTTCTGCTCCGAT 1343
QY 1158 TTTAATTTACTGATTTCCAGCAAGACCAAAATCATTTGATCAGATTATTTTAAAGTTTATC 1217
Db 1344 TTTAATTTACTGATTTCCAGCAAGACCAAAATCATTTGATCAGATTATTTTAAAGTTTATC 1403
QY 1218 CGTAGTTTTGATAAAAAGATTTTCTATTCCTTGGTTCTGTGAGAGAACCTTAATAAGTGCT 1277
Db 1404 CGTAGTTTTGATAAAAAGATTTTCTATTCCTTGGTTCTGTGAGAGAACCTTAATAAGTGCT 1463
QY 1278 ACTTTGCCATTAAGGCAGACTAGGGTTCAATGCTTTTTTACCCTTTAAAAAAAATTTGTAA 1337
Db 1464 ACTTTGCCATTAAGGCAGACTAGGGTTCAATGCTTTTTTACCCTTTAAAAAAAATTTGTAA 1523
QY 1338 AGTCTAGTTACTACTTTTCTTTGATTTTGCACGTTTGCAGCTTAGCCATCTCAAGCAACT 1397
Db 1524 AAGTCTAGTTACTACTTTTCTTTGATTTTGCACGTTTGCAGCTTAGCCATCTCAAGCAACT 1583
QY 1398 TTCGACGTTTGTACTAGCCATCTCAAGCAAGTTTTAATCAAGATCATCTCACGCTGATCAT 1457
Db 1584 TTCGACGTTTGTACTAGCCATCTCAAGCAAGTTTTAATCAAGATCATCTCACGCTGATCAT 1643
QY 1458 TGGATCCTACTCAACAAAAAGAGGTTGGTTCAGAGTACATTAAAGATTTCCTCTCCAA 1517
Db 1644 TGGATCCTACTCAACAAAAAGAGGTTGGTTCAGAGTACATTAAAGATTTCCTCTCCAA 1703
QY 1518 TTTTCAATAAATTTCTTCTCTCTTTAAAAAAAATTTAAAA 1563
Db 1704 TTTTCAATAAATTTCTTCTCTCTTTAAAAAAAATTTAAAA 1749
```

RESULT 5

```
AA162846
ID ID AAT62846 standard; DNA; 1749 BP.
XX
AC AAT62846;
XX
DT 25-MAR-2003 (updated)
DT 08-MAY-1997 (first entry)
XX
XX Human plasma carboxypeptidase B coding sequence.
DE
XX Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW plasma; plasminogen; ss.
XX
XX Homo sapiens.
FH Key Location/Qualifiers
FT CDS 41..1312
FT /*tag= a /product= Human PCPB
FT sig_peptide 41..106
FT /*tag= b
FT mat_peptide 107..1309
FT /*tag= c
FT misc_binding 134..177
FT /*tag= d
FT /bound_moeity= 46_bp_probe
XX
PN US5593674-A.
XX
XX 14-JAN-1997.
PD
XX 27-APR-1995; 95US-0430787.
XX
XX 01-FEB-1991; 91US-0649591.
PR
```



```
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
PR 27-APR-1995; 95US-0430787.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1997-099413/09.
XX P-PSDB; AAMI4733.
XX
XX Using human plasma carboxypeptidase B in blood coagulation - is
XX functionally related to carboxypeptidase A and pancreas
XX carboxypeptidase B
XX
XX Example 2; Column 37-42; 39pp; English.
XX
XX This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
XX PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX purified from human plasma or by transformed cell culture by
XX extraction using plasminogen bound to a solid phase.
XX (Updated on 25-MAR-2003 to correct PP field.)
XX
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
XX
XX Query Match 87.2%; Score 1372; DB 18; Length 1749;
XX Best Local Similarity 90.3%; Pred. No. 0;
XX Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
XX
XX 1 AGAAATTCGTGGGATGAAGCTTTCAGACCTTTCAGCTTCCTGATCCCATTTGCTCT 60
XX 24 AGAAATTCGTGGGATGAAGCTTTCAGACCTTTCAGCTTCCTGATCCCATTTGCTCT 83
XX
XX 61 TCTGTGAGCAGCATGCTTCGCGCTTCAGAGTCCGCAAGTTCAGCTCTCTAGAA 120
XX 84 TCTGTGAGCAGCATGCTTCGCGCTTCAGAGTCCGCAAGTTCAGCTCTCTAGAA 143
XX
XX 121 CCTCTAGGCAAGTTCAGATCTTACAGATCTTACTACACATATGAGATGTTCTCTGGC 180
XX 144 CCTCTAGGCAAGTTCAGATCTTACAGATCTTACTACACATATGAGATGTTCTCTGGC 203
XX
XX 181 AGCCGTAACAGCTGACCTTATTTGTAAGAAAAAACAAGTCCATTTTTTTGTAATGCAAT 240
XX 204 AGCCGTAACAGCTGACCTTATTTGTAAGAAAAAACAAGTCCATTTTTTTGTAATGCAAT 263
XX
XX 241 CTGATCTGCACATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTGC 300
XX 264 CTGATCTGCACATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTGC 323
XX
XX 301 TGGCAGACGTGGAAGATCTTATTTCAACAGCAGATTTTCAACGACACACAGTCAGCCCGGAG 360
XX 324 TGGCAGACGTGGAAGATCTTATTTCAACAGCAGATTTTCAACGACACACAGTCAGCCCGGAG 383
XX
XX 361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTTCTGGATAGAAT 420
XX 384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTTCTGGATAGAAT 443
XX
XX 421 TTATAACTGAGAGGATCTTGATATGCTTACAAAATTCACATGGAATCTCATTTTGAGA 480
XX 444 TTATAACTGAGAGGATCTTGATATGCTTACAAAATTCACATGGAATCTCATTTTGAGA 503
XX
XX 481 AGTACCCACTCTATGTTTTTAAAGTTTCTGGAAGAAACAAACAGCAAAATGCCATAT 540
XX 504 AGTACCCACTCTATGTTTTTAAAGTTTCTGGAAGAAACAAACAGCAAAATGCCATAT 563
XX
XX 541 GGATGACGTGGAATCCATGCCAGAGAAATGGAATCTCTCTGCTTTCTGCTTGGTTCA 600
XX 564 GGATGACGTGGAATCCATGCCAGAGAAATGGAATCTCTCTGCTTTCTGCTTGGTTCA 623
XX
XX 601 TAGGCAT----- 608
XX |||||
```

RESULT 6

AAQ41001

ID AAQ41001 standard; cDNA; 1749 BP.

XX AC

XX AAQ41001;

XX 24-AUG-1993 (first entry)

XX DT

XX Human plasma carboxypeptidase B gene.

XX DE

XX PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.

XX KW

XX XX

XX OS

XX Homo sapiens.

XX Key Location/Qualifiers

FH sig_peptide 41..106

FT /*tag= a

FT mat_peptide 107..1312

FT /*tag= b

FT /product= PCPB

FT misc_feature 134..177

FT /*tag= c

FT /function= probe

FT /note= "used to obtain full-length clones"

XX XX

XX US5206161-A.

XX XX

XX 27-APR-1993.

XX PD

XX 01-FEB-1991;

XX 91US-0649591.

XX 01-FEB-1991;

XX 91US-0649591.

XX XX

XX (GETH) GENENTECH INC.

XX PA

XX Drayna DT, Eaton DL;

XX PI

XX WFI; 1993-151724/18.

XX DR

XX P-PSDB; AAR36273.

XX DR

XX New human plasma carboxypeptidase B - used as haemostatic

XX PT

XX regulator for clotting blood, partic. for treating blood clotting

XX PT

XX disorders, e.g. haemophilia

XX PT

XX Disclosure; Fig 4; 40pp; English.

XX PS

XX Human plasma carboxypeptidase B was isolated from human plasma and

XX CC

XX partially sequenced. Oligonucleotide primers were designed based on

XX CC

XX the partial amino acid sequences. The primers were used in a PCR

XX CC

XX amplification to identify cDNA encoding PCPB from a human liver cDNA

XX CC

XX library. The PCR product was capable of encoding the first 37 amino

XX CC

XX acids of PCPB; a 46mer probe was used to obtain the full-length

XX CC

XX sequence which, although disclosed in the specification, is not

XX CC

XX claimed. PCPB inhibits the enzymatic conversion by tPA of

XX CC

XX plasminogen to plasmin in the presence of fibrinogen.

XX CC

XX Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

XX SQ

Query Match 87.1%; Score 1370.4; DB 14; Length 1749;

Best Local Similarity 90.2%; Pred. No. 0;

Matches 1557; Conservative 0; Mismatches 6; Indels 163; Gaps 2;

QY 1 AGAAAAATGCTGTGGATGAAGCTTTCAGCCCTTCAGTCCCTGTACCCATTGTTCTCT 60

Db 24 AGAAAAATGCTGTGGATGAAGCTTTCAGCCCTTCAGTCCCTGTACCCATTGTTCTCT 83

QY 61 TCTGTGAGCAGCATGCTTCGCGTTCCAGATGCCAGTCTAGCTCTTCCTAGAA 120

Db 84 TCTGTGAGCAGCATGCTTCGCGTTTCAGATGCCAGTCTAGCTCTTCCTAGAA 143

QY 121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180

Db	144	CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC	203
QY	181	AGCCGGTAACAGCTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTTTTTGTAAATGTCAT	240
Db	204	AGCCGGTAACAGCTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTTTTTGTAAATGTCAT	263
QY	241	CTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC	300
Db	264	CTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC	323
QY	301	TGSCAGACGTGGAAGATCTTATTTCAACAGCAGATTTTCAACAGCAGTCCAGTCCCGGAG	360
Db	324	TGSCAGACGTGGAAGATCTTATTTCAACAGCAGATTTTCAACAGCAGTCCAGTCCCGGAG	383
QY	361	CCTCCGCATCGTACTATGAACAGTATCACTACTAAATGAATCTATTTTGGATAGAAT	420
Db	384	CCTCCGCATCGTACTATGAACAGTATCACTACTAAATGAATCTATTTTGGATAGAAT	443
QY	421	TTATAACTGAGAGGCATCTGTATGCTTTACAAAAATCCACATTTGGATCTCTCAATTTGAGA	480
Db	444	TTATAACTGAGAGGCATCTGTATGCTTTACAAAAATCCACATTTGGATCTCTCAATTTGAGA	503
QY	481	AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAGAAACAAAACAGCCAAAAATGCCATAT	540
Db	504	AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAGAAACAAAACAGCCAAAAATGCCATAT	563
QY	541	GGATTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTCTGCTTCTGCTGTTCA	600
Db	564	GGATTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTCTGCTTCTGCTGTTCA	623
QY	601	TAGGCCAT-----	608
Db	624	TAGGCCATATACTCAATTTCTATGGGAATAATAGGGCAATATACCAATCTCTCGAGGCTTG	683
QY	609	-----A	609
Db	684	TGGATTTCTATGTTATGCGCGTGGTTAATCTGACGGTTATGACTACTCATGGAAGAA	743
QY	610	ATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAAACATCATTTGATCGGAACAGACC	669
Db	744	ATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAAACATCATTTGATCGGAACAGACC	803
QY	670	TGAATAGCAACTTTGCTCCAAACACTGTTGAGGAAGTGCATCCAGTTCCTCATGCT	729
Db	804	TGAATAGCAACTTTGCTCCAAACACTGTTGAGGAAGTGCATCCAGTTCCTCATGCT	863
QY	730	CGGAAACCTACTGTGGACTTTTATCTGAGTCAAGACCAAGTGAAGGCAAGTGGCTAGTT	789
Db	864	CGGAAACCTACTGTGGACTTTTATCTGAGTCAAGACCAAGTGAAGGCAAGTGGCTAGTT	923
QY	790	TCTTGAGAGAAATATCAACAGATTAAGCATATACATCAGCATGCAATTCATCTCCAGC	849
Db	924	TCTTGAGAGAAATATCAACAGATTAAGCATATACATCAGCATGCAATTCATCTCCAGC	983
QY	850	ATATAGTGTTCATATTCCTATACAGAGTAAAGCAAGACCATGAGGAATGTCTC	909
Db	984	ATATAGTGTTCATATTCCTATACAGAGTAAAGCAAGACCATGAGGAATGTCTC	1043
QY	910	TAGTAGCCAGTGAAGCAGTTCTGCTTATTTGACAAAACTAGTAAAAATACAGGTATACAC	969
Db	1044	TAGTAGCCAGTGAAGCAGTTCTGCTTATTTGAGAAAACTAGTAAAAATACAGGTATACAC	1103
QY	970	ATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGTGGGAGCATTTGATCTATG	1029
Db	1104	ATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGTGGGAGCATTTGATCTATG	1163
QY	1030	ATTTGGGCATCAATATTCGT-----	1050
Db	1164	ATTTGGGCATCAATATTCGTTTACAAATTTGAACTTCGAGATACGGGCACATACGGATTCT	1223
QY	1051	-----TTACATCAAAACCCCTGTGAGAAAGCTTTTGGCGTGTCTCTAAAA	1097

Db 1224 TGCTGCGGAGCGTTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTCTAAA 1283
Qy 1098 TAGCTTGGCATGTCATTAGGAATGTTTAAATGCCCTGATTTATCATCTTCTGCTCCGTAT 1157
Db 1284 TAGCTTGGCATGTCATTAGGAATGTTTAAATGCCCTGATTTATCATCTTCTGCTCCGTAT 1343
Qy 1158 TTTAAATTTACTGATTCAGCAAGCAACCAATCATTTGATATCATGATTTTAAAGTTTATC 1217
Db 1344 TTTAAATTTACTGATTCAGCAAGCAACCAATCATTTGATATCATGATTTTAAAGTTTATC 1403
Qy 1218 CGTAGTTTTCATAAAGATTTTCTATTCTCTGTTCTGTGTCAGAGAACCTTAATAAGTGCT 1277
Db 1404 CGTAGTTTTCATAAAGATTTTCTATTCTCTGTTCTGTGTCAGAGAACCTTAATAAGTGCT 1463
Qy 1278 ACTTGGCATTAAGGAGACTAGGTTTCATGCTTTTACCTTTTAAAAAAATTTGTA 1337
Db 1464 ACTTGGCATTAAGGAGACTAGGTTTCATGCTTTTACCTTTTAAAAAAATTTGTA 1523
Qy 1338 AAGCTAGTTACCTACTTTTCTTGTATTTTCAGCTTTGACCTAGCCTCAAGCAACT 1397
Db 1524 AAGCTAGTTACCTACTTTTCTTGTATTTTCAGCTTTGACCTAGCCTCAAGCAACT 1583
Qy 1398 TTCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACGCTGATCAT 1457
Db 1584 TTCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACGCTGATCAT 1643
Qy 1458 TGGATCTACTCAACAAAGAAAGGGTGGTGCAGAGTACATTAAGATTTCTGCTCCAA 1517
Db 1644 TGGATCTACTCAACAAAGAAAGGGTGGTGCAGAGTACATTAAGATTTCTGCTCCAA 1703
Qy 1518 TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAATTTAAAAA 1563
Db 1704 TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAATTTAAAAA 1749

RESULT 7

AAV74302
ID AAV74302 standard; cDNA; 1272 BP.
XX
AC AAV74302;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
XX
KW polymorphism detection; thrombotic disease; ds.
XX
OS Homo sapiens.
XX
PN W09855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI; 1999-045800/04.
DR
DR P-PSDB; AAW92270.
XX
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
XX
PS Example 1; Page 24; 35pp; English.
XX
CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant

CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
SQ Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

Query Match 58.4%; Score 918; DB 20; Length 1272;
Best Local Similarity 86.8%; Pred. No. 2.8e-211;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;
Qy 18 ATGAAGCTTTGCAGCCTTGCGAGTCTTGACCCATTGTTCTCTCTGTGAGCAGCATGTC 77
Db 1 ATGAAGCTTTGCAGCCTTGCGAGTCTTGACCCATTGTTCTCTCTGTGAGCAGCATGTC 60
Qy 78 TTGCGGTTCCAGAGTGGCCAAAGTTCTAGCTGCTTCTTAGAACCTCTAGGCAAGTTCAA 137
Db 61 TTGCGGTTTCAGAGTGGCCAAAGTTCTAGCTGCTTCTTAGAACCTCTAGGCAAGTTCAA 120
Qy 138 GTTCTACAGATCTTACTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGAC 197
Db 121 GTTCTACAGATCTTACTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGAC 180
Qy 198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCACTCTGATGTGCAATGTG 257
Db 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCACTCTGATGTGCAATGTG 240
Qy 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGCAGACGTGGAGAT 317
Db 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGCAGACGTGGAGAT 300
Qy 318 CTTATTCAACAGCAGATTTTCCAACGACACAGTCAGCCCGCCGCGCTCCGATCGTACTAT 377
Db 301 CTTATTCAACAGCAGATTTTCCAACGACACAGTCAGCCCGCCGCGCTCCGATCGTACTAT 360
Qy 378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTAGAGGCAT 437
Db 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTAGAGGCAT 420
Qy 438 CTTGATATGCTTTACAAAAATCCACATTTGGATGCTCTCATTTGAGAAAGTACCCACTATGTT 497
Db 421 CTTGATATGCTTTACAAAAATCCACATTTGGATGCTCTCATTTGAGAAAGTACCCACTATGTT 480
Qy 498 TTAAGGTTTCTGGAAGAAAGAAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
Db 481 TTAAGGTTTCTGGAAGAAAGAAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
Qy 558 CATGCCAGAGAATGGATCTCTCTCTGTTCTGTTGTTTCATAGGCCAT----- 608
Db 541 CATGCCAGAGAATGGATCTCTCTCTGTTCTGTTGTTTCATAGGCCATATAAATCAA 600
Qy 609 ----- 608
Db 601 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 660
Qy 609 -----AATCGAATGTGGAGAAAG 626
Db 661 CGGTTGGTTAATGTGGAACGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGGAGAAAG 720
Qy 627 AACCGTTCTTTTCTATGCGAACAATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
Db 721 AACCGTTCTTTTCTATGCGAACAATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGTC 780
Qy 687 TCAAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGTCTCGAAACCTACTCTGTGA 746
|||||

Db 781 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTGCTCGGAAACCTACTGTGGA 840
Qy 747 CTTTATCCTGAGTCAGAACAGAGTGAAGCGAGTGGCTAGTTCTTTGAGAAGAAATATC 806
Db 841. CTTTATCCTGAGTCAGAACAGAGTGAAGCGAGTGGCTAGTTCTTTGAGAAGAAATATC 900
Qy 807 AACCAGATTAAAGCATACATCAGCATGCATTCATCTACTCCAGCATATAGTGTTCATAT 866
Db 901 AACCAGATTAAAGCATACATCAGCATGCATTCATCTACTCCAGCATATAGTGTTCATAT 960
Qy 867 TCCTATACAGAGTAAAGCAAGAACAGCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 926
Db 961 TCCTATACAGAGTAAAGCAAGAACAGCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 1020
Qy 927 GTTCGTGCTATTGACAACTAGTAAATACAGGTATACATGAGTGGGCGCTCAGAA 986
Db 1021 GTTCGTGCTATTGAGAACTAGTAAATACAGGTATACATGAGTGGGCGCTCAGAA 1080
Qy 987 ACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAATAT 1046
Db 1081 ACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAATAT 1140
Qy 1047 TCGT-----TTAC 1054
Db 1141 TCGTTTACAATTGAACCTTCAGATACGGGCACATACGGAATCTTGTGCGGAGCGTTAC 1200
Qy 1055 ATCAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAATAGCTTGGCATGTCATT 1114
Db 1201 ATCAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAATAGCTTGGCATGTCATT 1260
Qy 1115 AGGAATGTTTAA 1126
Db 1261 AGGAATGTTTAA 1272

RESULT 8
AAF18005
ID AAF18005 standard; DNA; 1400 BP.
XX AAF18005;
XX
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
XX Human; lung cancer associated protein; neuroprotective; cytotostatic;
XX cardioactive; immunomodulatory; muscular active; vulnerable;
XX gastrointestinal; nephrotropic; antiinfective; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
XX
XX WO20005180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
XX P-PSDB; AAB58129.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -
XX

PS Claim 1; Page 507; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytotostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
Query Match 54.3%; Score 853.8; DB 21; Length 1400;
Best Local Similarity 93.3%; Pred. No. 8.5e-196;
Matches 934; Conservative 1; Mismatches 13; Indels 53; Gaps 2;
Qy 607 ATAAATCGAATGGGAGAAAGAACCGTTCTTATCGGAACATCATTCGATCGGACAG 666
Db 177 AGAATCGAATGGGAGAAAGAACCGTTCTTATCGGAACATCATTCGATCGGACAG 236
Qy 667 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAGGTGATCCAGTTCCTCAT 726
Db 237 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAGGTGATCCAGTTCCTCAT 296
Qy 727 GCTCGGAAACCTACTCTGTGGAATTTATCTGAGTCAGAACCCAGAGTGAAGGCGAGTCTA 786
Db 297 GCTCGGAAACCTACTCTGTGGAATTTATCTGAGTCAGAACCCAGAGTGAAGGCGAGTCTA 356
Qy 787 GTTCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGCTTACTACTCCC 846
Db 357 GTTCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGCTTACTACTCCC 416
Qy 847 AGCATATAGTGTTCCTATATTCCTATACAGAGTAAAGCAAAAGCAACCATGAGGAACTGT 906
Db 417 AGCATATAGTGTTCCTATATTCCTATACAGAGTAAAGCAAAAGCAACCATGAGGAACTGT 476
Qy 907 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAATAATACCAGGTATA 966
Db 477 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAATAATACCAGGTATA 536
Qy 967 CACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGATCT 1026
Db 537 CACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGATCT 596
Qy 1027 ATGATTTGGGCATCAAAATATTCTG----- 1050
Db 597 ATGATTTGGGCATCAAAATATTCTGTTTACAAATGGAATCTCGAGATACGGGCACATACGGAT 656
Qy 1051 -----TTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTA 1094
Db 657 TCTTGTGCGGAGCGTTATACCAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTA 716
Qy 1095 AAATAGCTTGGCATGTCTATAGGAATGTTTAATGCCCTGATTTTATCATCTGCTTCCG 1154
Db 717 AAATAGCTTGGCATGTCTATAGGAATGTTTAATGCCCTGATTTTATCATCTGCTTCCG 776
Qy 1155 TATTTTAAATTTACTGATTCAGCAAGCAAACTATTGTATCATGATTTATTTTAAAGTTT 1214
Db 777 TATTTTAAATTTACTGATTCAGCAAGCAAACTATTGTATCATGATTTATTTTAAAGTTT 836
Qy 1215 ATCCGTAGTTTGTGATAAAGATTTTCTATTCCTGTTCTGTGAGAAACCTAATAAGT 1274

Db	837	ATCCGTAGTCTTTTGATAAAAGATTTTCTATTCCTTGTTCTGTGCAGAGAACCTTAATAAGT	896
Qy	1275	GCTACTTTTGCCATTAAAGCAGACTAGGGTTCAATGCTCTTTTACCCCTTTAAAAAAAATTG	1334
Db	897	GCTACTTTTGCCATTAAAGCAGACTAGGGTTCAATGCTCTTTTACCCCTTT-AAAAAAAATTG	955
Qy	1335	TAAAGTCTAGTTACCTACTTTTTCTTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCA	1394
Db	956	TAAAGTCTAGTTACCTACTTTTTCTTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCA	1015
Qy	1395	ACTTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAAGATCATCTCAGCGTGA	1454
Db	1016	ACTTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAAGATCATCTCAGCGTGA	1075
Qy	1455	CATTGGATCCTACTCAACAAAGAGAGGTGGTCAGAAAGTACATTAAAGATTTCTGCTCC	1514
Db	1076	CATTGGATCCTACTCAACAAAGAGAGGTGGTCAGAAAGTACATTAAAGATTTCTGCTCC	1135
Qy	1515	AAATTTTCAATAAATTTCTTCTCTCCCTTTAAAAAAAATA	1555
Db	1136	AAATTTTCAATAAATTTCTGCTTGCTTTAGAAATACAA	1176

RESULT 9

ABX13671
ID ABX13671 standard: DNA: 55827 BP.

```
FT variation replace(12130,T)
FT /*tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT replace(17867..17868,GAA)
FT /*tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT replace(18243,G)
FT /*tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT replace(22450,C)
FT /*tag= aj
FT /standard_name= "Single nucleotide polymorphism"
FT replace(23003,T)
FT /*tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT replace(24055..24056,GAA)
FT /*tag= al
FT /standard_name= "Single nucleotide polymorphism"
FT replace(24132,C)
FT /*tag= am
FT /standard_name= "Single nucleotide polymorphism"
FT replace(25713,A)
FT /*tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28476,C)
FT /*tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT replace(29404,G)
FT /*tag= ap
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31047,T)
FT /*tag= aq
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31445,C)
FT /*tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31447,C)
FT /*tag= as
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31600,C)
FT /*tag= at
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31714,C)
FT /*tag= au
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31715,T)
FT /*tag= av
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32193,G)
FT /*tag= aw
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32341,G,A)
FT /*tag= ax
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32561,G)
FT /*tag= ay
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32600..35601,AAA)
FT /*tag= az
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32642,T)
FT /*tag= ba
FT /standard_name= "Single nucleotide polymorphism"
FT replace(32793..32794,TAG)
FT /*tag= bb
FT /standard_name= "Single nucleotide polymorphism"
FT replace(33071,A)
FT /*tag= bc
FT /standard_name= "Single nucleotide polymorphism"
FT replace(34721,T)
FT /*tag= bd
FT /standard_name= "Single nucleotide polymorphism"
FT replace(35304,A)

FT /*tag= be
FT /standard_name= "Single nucleotide polymorphism"
FT replace(35425,C)
FT /*tag= bf
FT /standard_name= "Single nucleotide polymorphism"
FT replace(36050,G)

Query Match 30.4%; Score 478.6; DB 25; Length 55827;
Best Local Similarity 98.0%; Pred. No. 7.6e-105;
Matches 495; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1051 TTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGGTGCTCTTAAATAGCTTGGCAATG 1110
Db 53334 TTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGGTGCTCTCTTAAATAGCTTGGCAATG 53393
QY 1111 CATTAGGAATGTTTAATGCCCTGATTTATCATCTGCTTCGCTATTTTAAATTTACTGA 1170
Db 53394 CATTAGGAATGTTTAATGCCCTGATTTTATCATCTGCTTCGCTATTTTAAATTTACTGA 53453
QY 1171 TTCAGCAAGACCAAAATCATTTGATCAGATTTATTTTAAGTTTATCCGTTAGTTTGGATA 1230
Db 53454 TTCAGCAAGACCAAAATCATTTGATCAGATTTATTTTAAGTTTATCCGTTAGTTTGGATA 53513
QY 1231 AAAGATTTTCCTATTCCTTGGTTCTGTGAGAGAACCTTAATAGTGTCTTTCGCAATTA 1290
Db 53514 AAAGATTTTCCTATTCCTTGGTTCTGTGAGAGAACCTTAATAGTGTCTTTCGCAATTA 53573
QY 1291 GGCAGACTAGGTTTCATGCTTTTACCCTTTAAAAAAATTTGTAAGTCTAGTTTACC 1350
Db 53574 GGCAGACTAGGTTTCATGCTTTTACCCTTTTAAAAAAATTTGTAAGTCTAGTTTACC 53632
QY 1351 TACTTTTCTTTCATTTTCGACGTTTGACTAGCACTCTCAAGCAACTTTTCGACGTTTGAC 1410
Db 53632 TACTTTTCTTTCATTTTCGACGTTTGACTAGCACTCTCAAGCAACTTTTCGACGTTTGAC 53692
QY 1411 TAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAAGCTGATCATTTGGATCCTACTCA 1470
Db 53692 TAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAAGCTGATCATTTGGATCCTACTCA 53752
QY 1471 ACAAAGGAGGTTGTCAGAGTACATTAAGATTTCTGCTCCAAATTTTCAATAAAT 1530
Db 53752 ACAAAGGAGGTTGTCAGAGTACATTAAGATTTCTGCTCCAAATTTTCAATAAAT 53812
QY 1531 TCTTCTTCTCTTTTAAAAAATAA 1555
Db 53812 TCTGTTGCTTTAGAAATACAA 53837

RESULT 10
ABX49430
ID ABX49430 standard; cDNA; 416 BP.
XX
AC ABX49430;
XX
DT 21-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #14595.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
OS Bos Taurus.
XX
PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-0960352.
XX
PR 12-JAN-1999; 99US-115707P.
XX
PR 11-JAN-2000; 2000US-0480902.
XX
```


Db 61 GTAGTTTGTAAAGAATTTTCTTATTCCTTGGTTCGTGTCAGAGAACCTTAATAAGTGCTA 120
QY 1279 CTTTGCCATTAAAGCAGACAGTAGGTTTCATGCTCTTTTACCTTTTAAAAAAAATTGTAAA 1338
Db 121 CTTTGCCATTAAAGCAGACAGTAGGTTTCATGCTCTTTTACCTTTTNNNNNNNNNTTGTAAA 180
QY 1339 AGTCTAGTTACCTACTTTTCTTTTGAATTTTCGAGTTTGTAGTCCATCTCAAGCAACTT 1398
Db 181 AGTCTAGTTACCTACTTTTCTTTTGAATTTTCGAGTTTGTAGTCCATCTCAAGCAACTT 240
QY 1399 TCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGATCATTT 1458
Db 241 TCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGATCATTT 300
QY 1459 GGATCCCTACTCAACAAAGGAGGTTGTCAGA 1491
Db 301 GGATCCCTACTCAACAAAGGAGGTTGTCAGA 333

RESULT 12

AAH87390
ID AAH87390 standard; DNA; 333 BP.

AC AAH87390;

DT 27-FEB-2002 (first entry)

DE Human single nucleotide polymorphism containing DNA sequence #2247.

XX Biallelic marker; polymorphism; human; disease; diagnosis; treatment;
KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
KW transgenic; single nucleotide polymorphism; SNP; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FH Variation replace(238,G)
FT /*Tag= a
FT /standard_name= "single nucleotide polymorphism"

XX WO9953095-A2.

XX 21-OCT-1999.

XX 30-MAR-1999; 99WO-US06893.

XX 09-APR-1998; 98US-0057871.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Lander ES, Wang D, Hudson T;

XX WPI; 1999-620443/53.

XX Polymorphic human genomic sequences and related allele-specific probes
PT and primers, useful for genetic analysis, e.g. diagnosis and monitoring
PT of disease.

PS Claim 1; Page 283; 330pp; English.

XX This invention describes novel human nucleic acid segments (I)
CC containing polymorphic sites. The polymucleotides of (I) are used for,
CC e.g. correlating disease polymorphisms (or disease susceptibility) or
CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,
CC inflammation, heart or central nervous system disorders; detecting
CC susceptibility to microbial infection; treating or preventing such
CC diseases; forensic analysis; gene therapy; paternity testing; mapping
CC genomic loci associated with phenotypic traits (and subsequent cloning
CC of the genes responsible); and the production of transgenic organisms.
CC Antibodies raised against (I) are useful as diagnostic and therapeutic
CC tools and in drug screening. AAH85144 - AAH87644 represent the human
CC DNA sequences containing biallelic polymorphic sites described in the

CC invention.

XX Sequence 333 BP; 90 A; 64 C; 53 G; 117 T; 9 other;
SQ

Query Match 20.6%; Score 324; DB 20; Length 333;
Best Local Similarity 97.3%; Pred. No. 2.5e-68;
Matches 324; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1159 TTAATTTACTGATTCAGCAAGACCAAAATCATTTGTATCAGATTTATTTAAAGTTTATCC 1218
Db 1 TTAATTTACTGATTCAGCAAGACCAAAATCATTTGTATCAGATTTATTTAAAGTTTATCC 60

QY 1219 GTAGTTTGTATTAAGATTTTCTTATTTCTTGTCTGTGTCAGAGAACCTTAATAAGTGCTA 1278
Db 61 GTAGTTTGTATTAAGATTTTCTTATTTCTTGTCTGTGTCAGAGAACCTTAATAAGTGCTA 120

QY 1279 CTTTGCCATTAAAGCAGACAGTAGGTTTCATGCTCTTTTACCTTTAAAAAAATTGTAAA 1338
Db 121 CTTTGCCATTAAAGCAGACAGTAGGTTTCATGCTCTTTTACCTTTNNNNNNNNNTTGTAAA 180

QY 1339 AGTCTAGTTACCTACTTTTCTTGTATTTTCGAGTTTGTAGTCCATCTCAGCAACTT 1398
Db 181 AGTCTAGTTACCTACTTTTCTTGTATTTTCGAGTTTGTAGTCCATCTCAGCAACTT 240

QY 1399 TCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGATCATTT 1458
Db 241 TCGAGCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGATCATTT 300

QY 1459 GGATCCCTACTCAACAAAGGAGGTTGTCAGA 1491
Db 301 GGATCCCTACTCAACAAAGGAGGTTGTCAGA 333

RESULT 13

ABX43260
ID ABX43260 standard; cDNA; 397 BP.

AC ABX43260;

DT 20-FEB-2003 (first entry)

DE Bovine EST associated with lactation/muscle/fat deposition #8425.

XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.

OS Bos Taurus.

PN US2002137139-A1.

PD 26-SEP-2002.

PF 24-SEP-2001; 2001US-0960352.

PR 12-JAN-1999; 99US-115707P.

PR 11-JAN-2000; 2000US-0480902.

XX (BYAT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

XX Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI; 2003-110599/10.

XX New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle

PS Claim 2; SEQ ID No 8425; 245pp; English.

Result No.	Score	Match	Query	Length	DB	ID	Description	
1	1469.8	93.4	1625	4	US-09-813-133A-1	Sequence 1,	Appli	
2	1372	87.2	1749	1	US-07-649-591B-2	Sequence 2,	Appli	
3	1372	87.2	1749	1	US-08-277-540-2	Sequence 2,	Appli	
4	1372	87.2	1749	1	US-08-430-787A-2	Sequence 2,	Appli	
5	918	58.4	1272	2	US-08-869-057-1	Sequence 1,	Appli	
6	478.6	30.4	55827	4	US-09-813-133A-3	Sequence 3,	Appli	
7	124	7.9	927	2	US-08-782-760-5	Sequence 5,	Appli	
8	134	7.9	927	5	PC7-US96-00995-5	Sequence 5,	Appli	
9	123.4	7.8	921	1	US-08-696-139-3	Sequence 3,	Appli	
10	123.4	7.8	1215	1	US-08-696-139-1	Sequence 1,	Appli	
11	109	6.5	1263	4	US-09-011-769A-38	Sequence 38,	Appli	
12	107.4	6.8	999	2	US-08-860-882A-67	Sequence 67,	Appli	
13	107.4	6.8	999	4	US-09-011-769A-50	Sequence 50,	Appli	
14	107.4	6.8	1053	2	US-08-860-882A-64	Sequence 64,	Appli	
15	107.4	6.8	1053	4	US-09-011-769A-46	Sequence 46,	Appli	
16	107.4	6.8	1263	2	US-08-860-882A-56	Sequence 56,	Appli	
17	107.4	6.8	1284	2	US-08-860-882A-71	Sequence 71,	Appli	
18	107.4	6.8	1284	4	US-09-011-769A-55	Sequence 55,	Appli	
19	104.2	6.6	1053	4	US-09-463-451-27	Sequence 27,	Appli	
20	104.2	6.6	1053	4	US-09-463-451-28	Sequence 28,	Appli	
21	104.2	6.6	1059	2	US-08-860-882A-74	Sequence 74,	Appli	
22	104.2	6.6	1059	2	US-08-860-882A-77	Sequence 77,	Appli	
23	104.2	6.6	1059	4	US-09-011-769A-59	Sequence 59,	Appli	
24	104.2	6.6	1059	4	US-09-011-769A-63	Sequence 63,	Appli	
25	101	6.4	1870	3	US-09-171-945-112	Sequence 112,	App	
26	101	6.4	2154	3	US-09-171-945-124	Sequence 124,	App	
27	92.8	5.9	1311	4	US-09-675-305-9	Sequence 9,	Appli	

Db 361 CTCGCTCCTGATCTATGAACAGATACCTCACTAAATGAAATCTATTCTTGGATAGAATT 420
Qy 422 TATAACTGAGGCGATCCTGATATGCTTACAAATCCATTTGGATCCTCATTTGAGAA 481
Db 421 TATAACTGAGGCGATCCTGATATGCTTACAAATCCATTTGGATCCTCATTTGAGAA 480
Qy 482 GTACCCACTCTATGTTTAAAGGTTCTGGAAGAAACAAACAGCCAAATGCGCATATG 541
Db 481 GTACCCACTCTATGTTTAAAGGTTCTGGAAGAAACAAACAGCCAAATGCGCATATG 540
Qy 542 GATTGACTGTGGAATCCATCCAGAGAAATGATCTCTCTGCTTTTCTGCTTGTGTTTCAT 601
Db 541 GATTGACTGTGGAATCCATCCAGAGAAATGATCTCTCTGCTTTTCTGCTTGTGTTTCAT 600
Qy 602 AGGCCATAATCGAATGCGGAGAAAGAACCGTTCTTCTATGCGAACAATCATTTGCATCGG 661
Db 601 AGGCCATAATCGAATGCGGAGAAAGAACCGTTCTTCTATGCGAACAATCATTTGCATCGG 660
Qy 662 AACAGACTGTAATAGCAACTTTGTTCTCCAAACACTGTTGTGGAAGGTGCATCCAGTTC 721
Db 661 AACAGACTGTAATAGCAACTTTGTTCTCCAAACACTGTTGTGGAAGGTGCATCCAGTTC 720
Qy 722 CTCATGCTCGGAAACCTACTGTGGACTTTATCTGAGTCAGAACCCAGAGTGAAGGCAAT 781
Db 721 CTCATGCTCGGAAACCTACTGTGGACTTTATCTGAGTCAGAACCCAGAGTGAAGGCAAT 780
Qy 782 GGCTAGTTCTTGGAAGAATAATCAACAGATTAAGCATACATCAGCATGATTCATTA 841
Db 781 GGCTAGTTCTTGGAAGAATAATCAACAGATTAAGCATACATCAGCATGATTCATTA 840
Qy 842 CTCACGATATAGTGTTCATATCTTATACAGGAAGTAAAGCAAGACCATGAGGA 901
Db 841 CTCACGATATAGTGTTCATATCTTATACAGGAAGTAAAGCAAGACCATGAGGA 900
Qy 902 ACTGCTCTAGTAGCCAGTGAAGCAGTTCGTCTATTGACAAACTAGTAAATAACAG 961
Db 901 ACTGCTCTAGTAGCCAGTGAAGCAGTTCGTCTATTGAGAAATAGTAAATAACAG 960
Qy 962 GTATACATAGGCGATGGCTCAGAACTTATACCTAGCTCCTGGAGTGGGACGATTG 1021
Db 961 GTATACATAGGCGATGGCTCAGAACTTATACCTAGCTCCTGGAGTGGGACGATTG 1020
Qy 1022 GATCTATGATTGGGCATCAAAATTCGT----- 1050
Db 1021 GATCTATGATTGGGCATCAAAATTCGT----- 1080
Qy 1051 -----TTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGT 1089
Db 1081 CGGATTCCTGTCGCGGAGGCTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGT 1140
Qy 1090 CTCTAAATAGCTGGCATGTCTAGGAATGTTTAAATGCCCTGTATTTATCATCTTGC 1149
Db 1141 CTCTAAATAGCTGGCATGTCTAGGAATGTTTAAATGCCCTGTATTTATCATCTTGC 1200
Qy 1150 TTCCGTAATTTTAAATTTACTGATCCAGACAGCAATCATTTGATCAGATTAATTTAA 1209
Db 1201 TTCCGTAATTTTAAATTTACTGATCCAGACAGCAATCATTTGATCAGATTAATTTAA 1260
Qy 1210 GTTTTATCCGTAGTTTGTATAAAGATTTTCTTATTCCTTGGTTCTGTCAGAGAACCTAA 1269
Db 1261 GTTTTATCCGTAGTTTGTATAAAGATTTTCTTATTCCTTGGTTCTGTCAGAGAACCTAA 1320
Qy 1270 TAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTGATGCTTTTACCCCTTTAAAAAA 1329
Db 1321 TAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTGATGCTTTTACCCCTTT-AAAAA 1379
Qy 1330 AATTGTAAGTCTAGTACTCTTTTCTTTGATTTTTCGACTTTGACTAGCCATCTC 1389
Db 1380 AATTGTAAGTCTAGTACTCTTTTCTTTGATTTTTCGACTTTGACTAGCCATCTC 1439
Qy 1390 AAGCAACTTTTCGAGTTTGTAGTCCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACG 1449
Db 1440 AAGCAAGTTTTCGAGTTTGTAGTCCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACG 1499

Qy 1450 CTGATCATTTGGATCCTACTCAACAAAGGAAGGTTGTCAGAAAGTACATTAAAGATTCT 1509
Db 1500 CTGATCATTTGGATCCTACTCAACAAAGGAAGGTTGTCAGAAAGTACATTAAAGATTCT 1559
Qy 1510 GTCTCCAAATTTTCAATAAATTTCTTCTTCTCTTTTAAAAAATTTTTTTTTTTTTTTT 1569
Db 1560 GTCTCCAAATTTTCAATAAATTTCTTCTTCTCTTTTAAAAAATTTTTTTTTTTTTTTT 1619
Qy 1570 AAA 1572
Db 1620 ATA 1622

RESULT 2

US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-07-649-591B-2

Query Match 87.2%; Score 1372; DB 1; Length 1749;

Best Local Similarity 90.3%; Pred. No. 0; Mismatches 5; Indels 163; Gaps 2;
Matches 1558; Conservative 0;

QY 1 AGAAAAATGCTGTGGGATGAAGCTTTGACAGCCTTGACGCTCTTGACCCATGTTCTCT 60
Db |||||
QY 24 AGAAAAATGCTGTGGGATGAAGCTTTGACAGCCTTGACGCTCTTGACCCATGTTCTCT 83
Db |||||
QY 61 TCTGTAGCAGCATGCTCTGCGCTTCAGAGTGCCCAAGTTCAGCTGCTCTTCCTAGAA 120
Db |||||
QY 84 TCTGTAGCAGCATGCTCTGCGCTTCAGAGTGCCCAAGTTCAGCTGCTCTTCCTAGAA 143
Db |||||
QY 121 CCTCTAGGCAAGTTCAGGCTTACAGATCTTACTCAACATATGAGATGTTCTCTGCG 180
Db |||||
QY 144 CCTCTAGGCAAGTTCAGGCTTACAGATCTTACTCAACATATGAGATGTTCTCTGCG 203
Db |||||
QY 181 AGCCGGTAACAGCTGACCTTATTTGTAAGAAAAACAAGTCCATTTTTTTGTAAGTGCAT 240
Db |||||
QY 204 AGCCGGTAACAGCTGACCTTATTTGTAAGAAAAACAAGTCCATTTTTTTGTAAGTGCAT 263
Db |||||
QY 241 CTGATGTGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTCTGCG 300
Db |||||
QY 264 CTGATGTGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTCTGCG 323
Db |||||
QY 301 TGGCAGAGTGGAGATCTTTATTCAACAGCAGATTTTCCACGACACAGTCAGCCCGGAG 360
Db |||||
QY 324 TGGCAGAGTGGAGATCTTTATTCAACAGCAGATTTTCCACGACACAGTCAGCCCGGAG 383
Db |||||
QY 361 CCTCCGATCGTACTATGACAGTATCACTCACTAAATGAAATCTATTCTTGATAGAAT 420
Db |||||
QY 384 CCTCCGATCGTACTATGACAGTATCACTCACTAAATGAAATCTATTCTTGATAGAAT 443
Db |||||
QY 421 TTATACTGAGAGCGATCTGTATATGCTTACAAAAATCCACATGATGATCTCAATTTGAGA 480
Db |||||
QY 444 TTATACTGAGAGCGATCTGTATATGCTTACAAAAATCCACATGATGATCTCAATTTGAGA 503
Db |||||
QY 481 AGTACCCACTATGTTTAAAGGTTCTGGAAGAAACAAACAGCCCAAAATGCCATAT 540
Db |||||
QY 504 AGTACCCACTATGTTTAAAGGTTCTGGAAGAAACAAACAGCCCAAAATGCCATAT 563
Db |||||
QY 541 GGATGACTGTGGAATCCATGCGAGAGATGGAATCTCTCTGCTTCTGCTGTGCTTCA 600
Db |||||
QY 564 GGATGACTGTGGAATCCATGCGAGAGATGGAATCTCTCTGCTTCTGCTGTGCTTCA 623
Db |||||
QY 601 TAGGCCAT----- 608
Db |||||
QY 624 TAGGCCATATACTCAATCTATGGATATAGGGCAATATACCAATCTCCTGAGGCTTG 683
Db |||||
QY 609 -----A 609
Db |||||
QY 684 TGGATTTCTATGTTATGCGCGTGGTTAATGTGGACGGTTATGACTACTCATGGAAGAAAGA 743
Db |||||
QY 610 ATCGAATGTGGAAGAAACCGTTCTTTCTATCGGAACATCATTTGATCGGAACAGACC 669
Db |||||
QY 744 ATCGAATGTGGAAGAAACCGTTCTTTCTATCGGAACATCATTTGATCGGAACAGACC 803
Db |||||
QY 670 TGAATAGCAACTTTGTCTCAAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCT 729
Db |||||
QY 804 TGAATAGCAACTTTGTCTTCAAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCT 863
Db |||||
QY 730 CGGAACCTACTGTGACTTTATCTGAGTCAGAACCCAGAGTGAAGGCGAGTGGCTAGTT 789
Db |||||
QY 864 CGGAACCTACTGTGACTTTATCTGAGTCAGAACCCAGAGTGAAGGCGAGTGGCTAGTT 923
Db |||||
QY 790 TCTTGAGAGAAATATCAACAGATTAAGCATATACATGATGATGATGATGATGATGATGATGAT 849
Db |||||
QY 924 TCTTGAGAGAAATATCAACAGATTAAGCATATACATGATGATGATGATGATGATGATGATGAT 983
Db |||||
QY 850 ATATAGTGTTCCTATATTCCTATACAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 909
Db |||||
QY 984 ATATAGTGTTCCTATATTCCTATACAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1043
Db |||||
QY 910 TAGTACGAGTGAAGGCTGTGCTTATGACAAAACTAGTAAAAATACAGGTATACAC 969
Db |||||

1044 TAGTACGAGTGAAGCAGTTCTGCTATTGTAGAGAAACTAGTAAAAATACCGGTATACAC 1103
QY |||||
970 ATGGCCATGGCTCAGAAACCTTTATACCTAGCTCTCTGGAGGTGGGACGATTTGGATCTATG 1029
Db |||||
1104 ATGGCCATGGCTCAGAAACCTTTATACCTAGCTCTCTGGAGGTGGGACGATTTGGATCTATG 1163
QY |||||
1030 ATTTGGGCATCAAAATATTCGT----- 1050
Db |||||
1164 ATTTGGGCATCAAAATATTCGTTTTACAAATTGAATCTTCAGATACGGGCACATACGGATTCT 1223
QY |||||
1051 -----TTACATCAAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
Db |||||
1224 TGTGCGCGAGGCTTACATCAAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283
QY |||||
1098 TAGCTTGGCATGTCAATTAGGAATGTTTAAATGCCCCTGATTTTATCATCTCTGCTTCGGTAT 1157
Db |||||
1284 TAGCTTGGCATGTCAATTAGGAATGTTTAAATGCCCCTGATTTTATCATCTCTGCTTCGGTAT 1343
QY |||||
1158 TTTAATTTACTGATTCACAGCAAGCAAAATCATTTGTATCAGATTTATTTTAAAGTTTATC 1217
Db |||||
1344 TTTAATTTACTGATTCACAGCAAGCAAAATCATTTGTATCAGATTTATTTTAAAGTTTATC 1403
QY |||||
1218 CGTAGTTTTGTATAAAGATTTTCTTATTCCTTGGTTCTGTCTAGAGAACCCTAATAAGTGCT 1277
Db |||||
1404 CGTAGTTTTGTATAAAGATTTTCTTATTCCTTGGTTCTGTCTAGAGAACCCTAATAAGTGCT 1463
QY |||||
1278 ACTTTGCCATTAAAGCAGACTAGGGTTCTATGCTTTTTTAAACCTTTTAAAAAAAATTTGTA 1337
Db |||||
1464 ACTTTGCCATTAAAGCAGACTAGGGTTCTATGCTTTTTTAAACCTTTTAAAAAAAATTTGTA 1523
QY |||||
1338 AAGTCTAGTTACTACTTTTCTTGTGATTTTGTGACTTTGCAAGTCTCAAGCAACT 1397
Db |||||
1524 AAGTCTAGTTACTACTTTTCTTGTGATTTTGTGACTTTGCAAGTCTCAAGCAACT 1583
QY |||||
1398 TTCGAGTTTGTACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACGCTGATCAT 1457
Db |||||
1584 TTCGAGTTTGTACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACGCTGATCAT 1643
QY |||||
1458 TGGATCTACTCAACAAAAAGAGGTGCTCAGAGTACATTAAGATTTCTGCTCCAAA 1517
Db |||||
1644 TGGATCTACTCAACAAAAAGAGGTGCTCAGAGTACATTAAGATTTCTGCTCCAAA 1703
QY |||||
1518 TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAAAAATTTTAA 1563
Db |||||
1704 TTTTCAATAAATTTCTTCTCTCTTTTAAAAAAAAAATTTTAA 1749

RESULT 3
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727

, FILING DATE: 15-DEC-1993
 , PRIOR APPLICATION DATA:
 , APPLICATION NUMBER: 07/959944
 , FILING DATE: 14-OCT-1992
 , PRIOR APPLICATION DATA:
 , APPLICATION NUMBER: 07/649591
 , FILING DATE: 01-FEB-91
 , ATTORNEY/AGENT INFORMATION:
 , NAME: Haasak, Janet E.
 , REGISTRATION NUMBER: 28,616
 , REFERENCE/DOCKET NUMBER: 689D1C1D1
 , TELECOMMUNICATION INFORMATION:
 , TELEPHONE: 415/225-1896
 , TELEFAX: 415/952-9881
 , TELEX: 910/371-7168
 , INFORMATION FOR SEQ ID NO: 2:
 , SEQUENCE CHARACTERISTICS:
 , LENGTH: 1749 bases
 , TYPE: nucleic acid
 , STRANDEDNESS: single
 , TOPOLOGY: linear
 , US-08-277-540-2

D	b	684	TGGAATTTCTATGTATTAGCCGGTGGTTAAATGTGACCGGTTATGACTACTCATGAAAAAAGA	743
Q	y	610	ATCGAATGTGGAGAAAGAACCGTTCCTTTCTATTCGGAACAATCAATGTCATCGGAACAGACC	669
D	b	744	ATCGAATGTGGAGAAAGAACCGTTCCTTTCTATTCGGAACAATCAATGTCATCGGAACAGACC	803
Q	y	670	TGAATAGCAAATTTTGCTCCTCAAAACACTGGGTGTGAGGAAGGTGCATCAGTTCTCATGCT	729
D	b	804	TGAATAGCAAATTTTGCTCCTCAAAACACTGGGTGTGAGGAAGGTGCATCAGTTCTCATGCT	863
Q	y	730	CGGAAACCTACTCTGGACCTTTATCTCTGAGTCAGAACCAAGGTGAAGGCAAGTGGCTAGTT	789
D	b	864	CGGAAACCTACTCTGGACCTTTATCTCTGAGTCAGAACCAAGGTGAAGGCAAGTGGCTAGTT	923
Q	y	790	TCITTGAGAAGAAATATCAACACAGATTAAAGCATACATCAGCATGCAATTCATCTCCCAGC	849
D	b	924	TCITTGAGAAGAAATATCAACACAGATTAAAGCATACATCAGCATGCAATTCATCTCCCAGC	983
Q	y	850	ATATAGTGTTCANATTTCTCTATACAGAAAGTAAAAGCAAGCAATGAGGAACACTGTCTC	909
D	b	984	ATATAGTGTTCANATTTCTCTATACAGAAAGTAAAAGCAAGCAATGAGGAACACTGTCTC	1043
Q	y	910	TAGTAGCCAGTGAAGCAGTTTCGTCTATTGCAAAAACTAGTAATAAATACCAGGTATACAC	969
D	b	1044	TAGTAGCCAGTGAAGCAGTTTCGTCTATTGCAAAAACTAGTAATAAATACCAGGTATACAC	1103
Q	y	970	ATGCCATGCTCAGAAACCTTATACCTAGCTCCTCGAGGTGGGACGATTTGGATCTATG	1029
D	b	1104	ATGCCATGCTCAGAAACCTTATACCTAGCTCCTCGAGGTGGGACGATTTGGATCTATG	1163
Q	y	1030	ATTGGGCATCAAAATATTCGT-----	1050
D	b	1164	ATTGGGCATCAAAATATTCGTTTTCAATTGAACTTCGAGATACGGGCACATACGGATTCCT	1223
Q	y	1051	-----TTACATCAAAACCCAACCTGTAGAGAAAGCTTTTGGCGCTGTCTCTAAAA	1097
D	b	1224	TGCTGCCGAGCGTTACATCAAAACCCACCTGTAGAGAAAGCTTTTGGCGCTGTCTCTAAAA	1283
Q	y	1098	TAGCTTGGCATGTCATTAGGAATGTTTTAATGCCCCCTGATTTTATCATCTCGCTTCGGTAT	1157
D	b	1284	TAGCTTGGCATGTCATTAGGAATGTTTTAATGCCCCCTGATTTTATCATCTCGCTTCGGTAT	1343
Q	y	1158	TTTAAATTTACTGATTCACAGCAAGACCAATCAATGTTATGATTTATTTTAAAGTTTATC	1217
D	b	1344	TTTAAATTTACTGATTCACAGCAAGACCAATCAATGTTATGATTTATTTTAAAGTTTATC	1403
Q	y	1218	CGTAGTTTTCATAAAAAGATTTTCTATTCCTTGTTCTGTFCAGAGAACCTTAATPAAGTCT	1277
D	b	1404	CGTAGTTTTCATAAAAAGATTTTCTATTCCTTGTTCTGTFCAGAGAACCTTAATPAAGTCT	1463
Q	y	1278	ACTTTGCGCATTAAGGCAGACTAGGGTTCATGTCCTTTTACCTTTTAAAAAATAATTTGTA	1337
D	b	1464	ACTTTGCGCATTAAGGCAGACTAGGGTTCATGTCCTTTTACCTTTTAAAAAATAATTTGTA	1523
Q	y	1338	AAGTCTAGTTTACCTACTCTTTTCTTTGATTTTTCAGACGTTTTCAGCTAGGCATCTCAAGCAACT	1397
D	b	1524	AAGTCTAGTTTACCTACTCTTTTCTTTGATTTTTCAGCGTTTTCAGCTAGGCATCTCAAGCAACT	1583
Q	y	1398	TTGCGCGTTTTCAGCTAGGCATCTCAAGCAAGTTTAAATCAAAAGATTCATCTCAGCGTGCAT	1457
D	b	1584	TTGCGCGTTTTCAGCTAGGCATCTCAAGCAAGTTTAAATCAAAAGATTCATCTCAGCGTGCAT	1643
Q	y	1458	TGGATCTCTACTCAACAAAAAGGAGGTGGTTCAGAGTACATTAAGATTTCTCTCTCCAAA	1517
D	b	1644	TGGATCTCTACTCAACAAAAAGGAGGTGGTTCAGAGTACATTAAGATTTCTCTCTCCAAA	1703
Q	y	1518	TTTTCAATAAATTTCTTTCTCTCTTTTAAAAAATAAAAAAAAAAAAAA	1563
D	b	1704	TTTTCAATAAATTTCTTTCTCTCTCTTTTAAAAAATAAAAAAAAAAAAAA	1749

Query Match	87.2%;	Score 1372;	DB 1;	Length 1749;
Best Local Similarity	90.3%;	Pred. No. 0;		
Matches 1558;	Conservative 0;	Mismatches 5;	Indels 163;	Gaps 2;
QY 1	AGAAAAATGCTGTGGGATGAAGCTTTGCAGCCCTTGCACTCCTGTGTACCCATTGTTCTCT	60		
DB 24	AGAAAAATGCTGTGGGATGAAGCTTTGCAGCCCTTGCACTCCTGTGTACCCATTGTTCTCT	83		
QY 61	TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAGTGTCTAGCTGCTCTCTCCAGAA	120		
DB 84	TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAGTGTCTAGCTGCTCTCTCCAGAA	143		
QY 121	CCTTAGGCAAGTTCAGATTTCTACAGAACTTTACTACAACATATGAGATGTGTTCTCTGCG	180		
DB 144	CCTTAGGCAAGTTCAGATTTCTACAGAACTTTACTACAACATATGAGATGTGTTCTCTGCG	203		
QY 181	AGCCGGTAAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTGTAAATGCAT	240		
DB 204	AGCCGGTAAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTGTAAATGCAT	263		
QY 241	CTGATGTGCACAAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTGTCTTGC	300		
DB 264	CTGATGTGCACAAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTGTCTTGC	323		

Db 1404 CGTAGTTTGTATAAAGATTTTCCTATTCTTGGTCTGTGTCAGAGAACCTAATAAGTGCT 1463
Qy 1278 ACTTGGCATTAAAGCAGACTAGGTTTCATGTCCTTTTACCCCTTAAAGAAATTTGTAA 1337
Db 1464 ACTTGGCATTAAAGCAGACTAGGTTTCATGTCCTTTTACCCCTTAAAGAAATTTGTAA 1523
Qy 1338 AAGTCTAGTTACCTACTTTTCTTTGATTTTTCGACGTTTTCGACCTAGCCATCTCAAGCAACT 1397
Db 1524 AAGTCTAGTTACCTACTTTTCTTTGATTTTTCGACGTTTTCGACCTAGCCATCTCAAGCAACT 1583
Qy 1398 TTGACCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACGCTGATCAT 1457
Db 1584 TTGACCTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCACGCTGATCAT 1643
Qy 1458 TGGATCTACTCAACAAAGGAAGGGTTCAGAGTACATTAAGATTTTCTGCTCAAAA 1517
Db 1644 TGGATCTACTCAACAAAGGAAGGGTTCAGAGTACATTAAGATTTTCTGCTCAAAA 1703
Qy 1518 TTTTCAATAAATTTCTTCTCTCTTTTAAAGAAATTTTAAAGAAATTTTCTGCTCAAAA 1563
Db 1704 TTTTCAATAAATTTCTTCTCTCTTTTAAAGAAATTTTAAAGAAATTTTCTGCTCAAAA 1749

RESULT 5

US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morber, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Washtien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-5411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PUBLICATION INFORMATION:
; AUTHORS: Eaton, Dan L.
; AUTHORS: Malloy, Beth E.
; AUTHORS: Tsai, Siao P.
; AUTHORS: Henzel, William
; AUTHORS: Drayna, Dennis
; TITLE: Isolation, Molecular Cloning, and Partial
; TITLE: Characterization of a No. 5985562el Carboxypeptidase B
; TITLE: from Human Plasma

; JOURNAL: J. Biol. Chem.

; VOLUME: 266

; ISSUE: 32

; PAGES: 21833-21838

; DATE: No. 5985562 15-1991

; US-08-869-057-1

Query Match 58.4%; Score 918; DB 2; Length 1272;
Best Local Similarity 86.8%; Pred. No. 5.6e-213;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

Qy 18 ATGAAGCTTTGACGCTTTCAGGCTTCCTTACCCATTTGTTCTTCTGTGTGAGCAGCATGTC 77
Db 1 ATGAAGCTTTGACGCTTTCAGGCTTCCTTACCCATTTGTTCTTCTGTGTGAGCAGCATGTC 60
Qy 78 TTGCGGTTCCAGAGTGGCCAGTTCTAGCTGCTTCTTAGAACCTCTAGGCAAGTTCAA 137
Db 61 TTGCGGTTTCAGAGTGGCCAGTTCTAGCTGCTTCTTAGAACCTCTAGGCAAGTTCAA 120
Qy 138 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGAC 197
Db 121 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGAC 180
Qy 198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAATGTG 257
Db 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAATGTG 240
Qy 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGCACTGTTCTTGTGGCAGACGTGGAGAT 317
Db 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCACTGTTCTTGTGGCAGACGTGGAGAT 300
Qy 318 CTTATTCAACAGCAGATTTTCCACGACAGCAGTCAGCGCCCGAGCCCTCCGATCGTACTAT 377
Db 301 CTTATTCAACAGCAGATTTTCCACGACAGCAGTCAGCGCCCGAGCCCTCCGATCGTACTAT 360
Qy 378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 437
Db 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 420
Qy 438 CTTGATATGCTTACAAAATCCACATTTGATTCCTCATTTTGAAGTACCCACTCTATGTT 497
Db 421 CTTGATATGCTTACAAAATCCACATTTGATTCCTCATTTTGAAGTACCCACTCTATGTT 480
Qy 498 TTAAGGTTTCTGGAAGAAAGCAACAGCAAAATGCCATATGGATTCGATCTGGAATC 557
Db 481 TTAAGGTTTCTGGAAGAAAGCAACAGCAAAATGCCATATGGATTCGATTCGATTCGGAATC 540
Qy 558 CATGCCAGAGATGGATCTCTCTGCTTTTCTGCTTGTGTTTCATAGGCCAT----- 608
Db 541 CATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATATAACTCAA 600
Qy 609 ----- 608
Db 601 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATG 660
Qy 609 -----AATCGAATGTGGAAG 626
Db 661 CCGTGGTTAATGTGGAAGGTTTATGACTACTATGGAAGAAAGAAATCGAATGTGGAAG 720
Qy 627 AACCGTTCTTTCTATGCGAACCAATCATTTGTCATCGGAACAGACCTGAATAGCAACTTTGTC 686
Db 721 AACCGTTCTTTCTATGCGAACCAATCATTTGTCATCGGAACAGACCTGAATAGCAACTTTGTC 780
Qy 687 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTATGTCGGAACACTCTAGTGGGA 746
Db 781 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTATGTCGGAACACTCTAGTGGGA 840
Qy 747 CTTTATCCTGAGTCAGAACCAAGGTCGAGGTCGCTAGTTCTTCTGGAAGCAATATC 806
Db 841 CTTTATCCTGAGTCAGAACCAAGGTCGAGGTCGCTAGTTCTTCTGGAAGCAATATC 900
Qy 807 AACGAGTTAAAGCATACATCAGCATGCTTCACTCTCCAGCATATAGTGTTCATAT 866

Db 362 GAATGTGGAGAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTGGGTGTAGACCCCA 421
Qy 673 ATAGCAACTTTGTCTCAAAACACTGGGTGAGGAAGTGCATCCAGTTCCTCATGCTCGG 732
Db 422 ACAGGAATTTTAATGC---TGGCTGGGTGGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG 478
Qy 733 AAACCTACTGTGGGACTTTATCTCTGAGTCAGAACAGAGTGAAGGCGAGTGGCTAGTTTCT 792
Db 479 AAACCTACTGTGGGAGCCAGCCAGAGTCTGAAGAAGAGACAAAGGCCCTGGCAGATTTCA 538
Qy 793 TGAGAAGAAATATCAACAGATTAAGATACATCAGCATGCTATCATCTCCAGCATGA 852
Db 539 TCCGCAACAACTCTCCACCATCAAGGCTACCTGACCATCCACTCATACTCAGATGA 598
Qy 853 TAGGTGTTCCATATCTCTATACGAAGTAAAGCAAGACCATGAGGAAGTCTCTCTAG 912
Db 599 TGCTCTACCTTACTCTCTATGACTACAAAGTGCCTGAGAACTATGAGGAATGAATGCC 658
Qy 913 TAGCCAGTGAAGCAGTTTCGTCTATTGACAAACACTAGTAAATAATACCAAGGTATACACATG 972
Db 659 TGGTGAAGGTGGCGAAGAGGCTTGCC---ACTCTGCATGGCACCAGTACACATAG 715
Qy 973 GCCATGCTCAGAACTTATACCTCTAGTCTGAGGTGGGAGCATGATGATCTATGATT 1032
Db 716 GCCCAGGAGCTACAAACATCTATCTGCTGCTGGGGATCTGACGACTGGTCTTATGATC 775
Qy 1033 TGGGCATCAATATCTCTTACATCAACCCACTGTAGAGAAGCTTTTCCGCTGTCTC 1092
Db 776 AGGAATCAAAATATCTCTTACCTTTGA-ACTCCGGGATACAGGCTTCTTTGGCTTTCTC 834

RESULT 8

PCT-US96-00995-5
; Sequence 5, Application PC/TUS9600995
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00995
; FILING DATE: 25-JAN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 1..927
PCT-US96-00995-5
Query Match 7.9%; Score 124; DB 5; Length 927;
Best Local Similarity 57.9%; Pred. No. 6.1e-21;
Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;
Qy 613 GAATGTGGAGAAAACCCGCTTCTTATGCGAACAATCATTCATCGGAACAGACCTGA 672
Db 362 GAATGTGGAGAAAACCCGCTCTACTATGGTGGAAAGTTCCTGCTGGGTGTAGACCCCA 421
Qy 673 ATAGCAACTTTGTCTCAAAACACTGGGTGAGGAAGTGCATCCAGTTCCTCATGCTCGG 732
Db 422 ACAGGAATTTTAATGC---TGGCTGGGTGGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG 478
Qy 733 AAACCTACTGTGAGCTTTATCTCTGAGTCAGAACAGAGTGAAGGCGAGTGGCTAGTTTCT 792
Db 479 AAACCTACTGTGAGCAGCCAGAGTCTGAAAAGAGACAAAGGCCCTGGCAGATTTCA 538
Qy 793 TGAGAAGAAATATCAACAGATTTAAAGCATATCATCAGCATGCAATTCATCTCCAGCATGA 852
Db 539 TCCGCAACAACTCTCCACCATCAAGGCTACCTGACCATCCACTCATACTCAGATGA 598
Qy 853 TAGGTGTTCCATATCTCTTATACGAAAGTAAAGCAAGACCATGAGGAAGTCTCTCTAG 912
Db 599 TGCTCTACCTTACTCTTATGACTACAAAGTGCCTGAGAACTATGAGGAATGAATGCC 658
Qy 913 TAGCCAGTGAAGCAGTTTCGTCTATTGACAAACACTAGTAAATAATACCAAGGTATACACATG 972
Db 659 TGGTGAAGGTGGCGAAGAGGCTTTGCC---ACTCTGCATGGCACCAGTACACATAG 715
Qy 973 GCCATGCTCAGAAACCTTATACCTCTAGTCTGAGGTGGGAGCATGATGATCTATGATT 1032
Db 716 GCCCAGGAGCTACAAACATCTATCTGCTGCTGGGGATCTGACGACTGGTCTTATGATC 775
Qy 1033 TGGGCATCAATATCTCTTACATCAACCCACTGTAGAGAAGCTTTTCCGCTGTCTC 1092
Db 776 AGGAATCAAAATATCTCTTACCTTTGA-ACTCCGGGATACAGGCTTCTTTGGCTTTCTC 834

RESULT 9

US-08-696-139-3
; Sequence 3, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersherberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993

Db 1029 CTATCCTGCTGCTGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAAATTCCTT 1088
Qy 1052 TACAT 1056
Db 1089 CACCT 1093

RESULT 11
US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011.769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995

INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1263 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid
; SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-011-769A-38

Query Match 6.9%; Score 109; DB 4; Length 1263;
Best Local Similarity 54.4%; Pred. No. 2.8e-17;
Matches 264; Conservative 0; Mismatches 215; Indels 6; Gaps 2;

Qy 572 GATCTCTCTGCTTCTGCTGTGTTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
Db 645 GCTCAATATTGATGGCTACATCTACACTGACCACAGCGGATTTGGAGAAAGATCG 704
Qy 632 TTCTTTTCATGCGAACAAATCATTCGATCGGACAGACTGAATAGCACTTTGTCTCAA 691
Db 705 CTCCACCCACTACTGGATCTAGCTGATGGACAGACCCCAACAGAAATTTGATGC--- 761
Qy 692 ACATGCTGTGAGGAAGTGCATCAGTTCTCATGCTCGAATCCTACTGTGACTTAA 751
Db 762 TGGTTGGTGAATTTGGAGGCTCTCGAAACCCCTGTGATGAATCTACTGTGACCTGC 821
Qy 752 TCCTGAGTCAGAACCAAGGAAGGAGGAGTGGCTAGTTCTTGAGAAAGAAATATCAACCA 811

Db 822 CGCAGAGTCTGAAAGAAACCAAGGCCCTGGCTGATTTTCATCGCAACAAACTCTCTTC 881
Qy 812 GATTAAGCATACATCAGCATGCAATTCATATCCAGCATATAGTGTTCATATTCCTA 871
Db 882 CATCAAGGCATATCTGACAAATCCACTCGTACTCCCAATGATGATCACTCCCTTACTCATA 941
Qy 872 TACACGAAGTAAAGCAAGACCATGAGGAACCTGTCTCTAGTAGCGAGTGAAGCAGTTCG 931
Db 942 TGCTTACAAACTCGGTGAGAACATGCTGAGTTGAATGCCCTGCTAAAGCTACTGTGAA 1001
Qy 932 TGCTATTGACAAACTAGTAAATAATACAGGTATACACATGCGCATGGCTCAGAAACCTT 991
Db 1002 AGAATTTGCC---TCACTGACGGCACCAGTACACATATGGCCCGGAGCTACAAACAAT 1058
Qy 992 ATACCTAGCTCTGGAGGTGGGACCAATGGATCTATGATTTGGGCATCAATATTCGTT 1051
Db 1059 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 1118
Qy 1052 TACAT 1056
Db 1119 CACCT 1123

RESULT 12
US-08-860-882A-67
; Sequence 67, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUM
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 999 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-860-882A-67

Query Match 6.8%; Score 107.4; DB 2; Length 999;
Best Local Similarity 54.2%; Pred. No. 6.4e-17;
Matches 263; Conservative 0; Mismatches 216; Indels 6; Gaps 2;
QY 572 GATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
DB 387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGCGGATTTTGGAGAAAGACTCG 446
QY 632 TTCTTTCTATGCGAACAATCATTTGCTGGAACAGCCTGAATAGCAACTTTGTCCTCAA 691
DB 447 CTCACCCCATCTGCAATCCACTCGTACTCCCAATGATGATCTACCTTACTCAT 683
QY 692 ACACGAGTAAAGCAACAGCAGCAGTCTCTAGTAGCCAGTGAAGAGTTCG 931
DB 684 TGCTTACAACTCGGTGAGAACATGCTGATGTTGAATGCCCTGCTAAAGCTACTGTGAA 743
QY 932 TGCTTATGACAACTAGTAAATACCAATGAGTGGCTGATGATTTGGGCAATCAATATTCGTT 991
DB 744 AGAATTTGCC---TCACCTGACGCAACAGTACATATGCGCAACAACTCTCTTC 623
QY 992 ATACTAGCTCCTGGAGTGGGACGATGATGATGATTTGGGCAATCAATATTCGTT 1051
DB 801 CTATCTGCTGCTGGGGCTCTGACGACTGGCTTATGACCAAGGATCAGATATTCCTT 860
QY 1052 TACAT 1056
DB 861 CACCT 865

RESULT 13

US-09-011-769A-50
; Sequence 50, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:

APPLICANT: SLATER, Anthony M.

BLAKEY, David C.

DAVIES, David H.

HENHAM, John F.

HENNEQUIN, Laurent F.A.

MARSHAM, Peter R.

DOWELL, Robert I.

TITLE OF INVENTION: Chemical Compounds

NUMBER OF SEQUENCES: 87

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pillsbury Madison & Sutro, LLP

STREET: 1100 New York Ave., N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 Mb disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/011,769A

FILING DATE: 13-Feb-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB96/01975

FILING DATE: 13-AUG-1996
APPLICATION NUMBER: GB 9612295.7
FILING DATE: 12-JUN-1996
APPLICATION NUMBER: GB 9611019.2
FILING DATE: 25-MAY-1996
APPLICATION NUMBER: GB 9516810.0
FILING DATE: 16-AUG-1995

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 999 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

FEATURE:

NAME/KEY: CDS

LOCATION: 1..987

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 67..987

SEQUENCE DESCRIPTION: SEQ ID NO: 50:

US-09-011-769A-50

Query Match 6.8%; Score 107.4; DB 4; Length 999;

Best Local Similarity 54.2%; Pred. No. 6.4e-17;

Matches 263; Conservative 0; Mismatches 216; Indels 6; Gaps 2;

QY 572 GATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631

DB 387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGCGGATTTTGGAGAAAGACTCG 446

QY 632 TTCTTTCTATGCGAACAATCATTTGCTGGAACAGCCTGAATAGCAACTTTGTCCTCAA 691

DB 447 CTCACCCCATCTGCAATCCACTCGTACTCCCAATGATGATCTACCTTACTCAT 683

QY 692 ACACGAGTAAAGCAACAGCAGCAGTCTCTAGTAGCCAGTGAAGAGTTCG 931

DB 684 TGCTTACAACTCGGTGAGAACATGCTGATGTTGAATGCCCTGCTAAAGCTACTGTGAA 743

QY 932 TGCTTATGACAACTAGTAAATACCAATGAGTGGCTGATGATTTGGGCAATCAATATTCGTT 991

DB 744 AGAATTTGCC---TCACCTGACGCAACAGTACATATGCGCAACAACTCTCTTC 623

QY 992 ATACTAGCTCCTGGAGTGGGACGATGATGATGATTTGGGCAATCAATATTCGTT 1051

DB 801 CTATCTGCTGCTGGGGCTCTGACGACTGGCTTATGACCAAGGATCAGATATTCCTT 860

QY 1052 TACAT 1056

DB 861 CACCT 865

RESULT 14

US-08-860-882A-64

; Sequence 64, Application US/08860882A

; Patent No. 5985281

; GENERAL INFORMATION:

APPLICANT: TAYLORSON, CHRISTOPHER JOHN

APPLICANT: EGGELTE, HENDRIKUS JOHANNES

APPLICANT: TARRAGONA-FIOL, ANTONIO

APPLICANT: RABIN, BRIAN ROBERT

	APPLICANT:	BOYLE, FRANCIS THOMAS	
	APPLICANT:	HENNAM, JOHN FREDERICK	
	APPLICANT:	BLAKELY, DAVID CHARLES	
	APPLICANT:	MARSHAM, PETER ROBERT	
	APPLICANT:	HEATON, DAVID WILLIAM	
	APPLICANT:	DAVIES, DAVID HUM	
	TITLE OF INVENTION:	CHEMICAL COMPOUNDS	
	NUMBER OF SEQUENCES:	77	
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE:	PILLSBURY, MADISON & SUTRO	
	STREET:	1100 NEW YORK AVENUE, N.W.	
	CITY:	WASHINGTON	
	STATE:	D.C.	
	COUNTRY:	USA	
	ZIP:	20005	
	COMPUTER READABLE FORM:		
	COMPUTER:	IBM compatible	
	OPERATING SYSTEM:	PC-DOS/MS-DOS	
	SOFTWARE:	Patent In Release #1.0, Version #1.25	
	CURRENT APPLICATION DATA:		
	APPLICATION NUMBER:	US/08/860,882A	
	FILING DATE:	JUNE 23, 1997	
	CLASSIFICATION:	424	
	ATTORNEY/AGENT INFORMATION:		
	NAME:	DONALD J. BIRD	
	REGISTRATION NUMBER:	25,323	
	REFERENCE/DOCKET NUMBER:	9901/238653	
	TELEPHONE:	(202) 861-3027	
	TELEFAX:	(202) 822-0944	
	TELEX:	6174627 CUSH	
	INFORMATION FOR SEQ ID NO:	64:	
	SEQUENCE CHARACTERISTICS:		
	LENGTH:	1053 bases	
	TYPE:	nucleic acid	
	STRANDEDNESS:	single	
	TOPOLOGY:	linear	
	US-08-860-882A-64		
	Query Match	6.8%; Score 107.4; DB 2; Length 1053;	
	Best Local Similarity	54.2%; Pred. No. 6.5e-17;	
	Matches	263; Conservative 0; Mismatches 216; Indels 6; Gaps 2;	
QY	572	GATCTCTCCTGCTTCTGGTTCATAGGCCATAATCGAATGTGGAGAAGACCG	631
DB	387	GCTCAATATTGATGGCTACTACCTACCTGGACGAAGCCGATTTTGAGAAAGACTCG	446
QY	632	TTCTTTTCTATGCGAACAAATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCAA	691
DB	447	CTCCACCCTACTGGATCTAGCTGATGGCACAGACCCCAACAGAAATTTTGATGC---	503
QY	692	ACACTGGTGTAGGAAGTGGATCAGTTCTCATGCTCGGAACCTACTGTGACACTTA	751
DB	504	TGGTTGGTGTGAATTTGGAGCCCTCGAAAACCCCTGTGATGAACTTACTGTGGACCTGC	563
QY	752	TCTGTAGTCAGAACCAAGTAGTGAAGCAGTGGCTAGTTTCTTGAGAAATATCAACCA	811
DB	564	CGCAGAGTCTGAAGAGGAGACCAAGGCCCTGGCTGATTTATCCGCAACAACCTCTCTTC	623
QY	812	GATTAAGCATACATCAGCATGCTTATCTACTCTCCAGCATATAGTGTTCCTATATCTTA	871
DB	624	CATCAAGGCATATCTGACAACTCCACTCGTACTCCCCAAATGATGATCTACCTTACTATA	683
QY	872	TACACGAAGTAAAGCAAGACCATGAGGAATGCTCTCTAGTAGCAGTGAAGCAGTTTCG	931
DB	684	TGCTTACAAACTCGGTGAGAACAAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA	743
QY	932	TGCTATTGACAAACTAGTAAAAATACCAAGGTATACATGCGCATGCCATGCTCAGAAACCTT	991
DB	744	AGAACTTGCC---TACTCGCAGGACCAAGTACACATATGCGCGGAGGCTACAACAAT	800
QY	992	ATACCTAGTCTCTGGAGGTGGGAGCAATTGGATCTATGATTTGGGCATCAAAATATTGCTT	1051

Db	447	CTCCACCATACTGGATCTAGCTGCTGTCATTTGGCACAGACCCCAACAGAAATTTTGATGC---	503
Qy	692	ACACTGGTGTGAGGAAGGTGCTCATTCAGTTCTCTCATGCTCGGAAACCTTACTGTGGACTTTA	751
Db	504	TGGTGTGTGTGAATTTGGAGCTCTCGAAACCCCTGTGATGAAACTTTACTGTGGACCTGC	563
Qy	752	TCCTGAGTCAGAACCAAGTGAAGCAGTGGCTAGTTCTTGAGAGAAATATCAACCA	811
Db	564	CGCAGAGTCTGAAGAGAGACCAGGCCCTGGCTGATTTTCATCCGCAACAACTCTCTTC	623
Qy	812	GATTAAAGCATACATCAGCATGCAATTCATATCCAGCATATAGTGTTCATATTCCTA	871
Db	624	CATCAAGGCATATCTGACATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA	683
Qy	872	TACACGAAGTAAAGCAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTGG	931
Db	684	TGCTTACAAACTCGGTGAGAACCAATGCTGAGTTGAATGCCCTGAAAGCTACTGTGAA	743
Qy	932	TGCTATTGCAAAACTAGTAAATAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT	991
Db	744	AGAACTTGCC---TCACTGCAACGACCAAGTACATATGGCCCGGAGCTACAAAT	800
Qy	992	ATACCTAGCTCTCTGGAGGTGGGACGATTTGGATCTATGATTTGGGCATCAAAATATTCGTT	1051
Db	801	CTATCCTGCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT	860
Qy	1052	TACAT	1056
Db	861	CACCT	865

Search completed: January 1, 2004, 15:40:55
Job time : 118.353 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 13:56:00 ; Search time 541.923 Seconds
(without alignments)
10046.789 Million cell updates/sec

Title: US-09-980-881A-1

Perfect score: 1573
Sequence: 1 agaaattgctgttgggatg.....aaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

```
Database :
...
1: /cgm2_6/pdata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgm2_6/pdata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgm2_6/pdata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgm2_6/pdata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgm2_6/pdata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgm2_6/pdata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgm2_6/pdata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgm2_6/pdata/2/pubpna/US09_PUBCOMB.seq.*
9: /cgm2_6/pdata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgm2_6/pdata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgm2_6/pdata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgm2_6/pdata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgm2_6/pdata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgm2_6/pdata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgm2_6/pdata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgm2_6/pdata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgm2_6/pdata/2/pubpna/US60_NEW_PUB.seq.*
18: /cgm2_6/pdata/2/pubpna/US60_PUBCOMB.seq.*
...

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1372	87.2	1728	10	US-09-880-107-2396		Sequence 2396, Appl
2	858.8	54.6	1272	13	US-10-379-836-1		Sequence 1, Appl
3	853.8	54.3	1400	9	US-09-925-302-34		Sequence 24, Appl
4	331.2	21.1	416	10	US-09-960-352-14595		Sequence 14595, A
5	286.8	18.2	397	10	US-09-960-352-8425		Sequence 8425, Ap
6	137.2	8.7	431	10	US-09-917-800A-468		Sequence 468, App
7	131	8.3	1547	13	US-10-027-632-265133		Sequence 265133, S
8	131	8.3	1547	13	US-10-027-632-265134		Sequence 265134, S
9	131	8.3	1547	13	US-10-027-632-265135		Sequence 265135, S
10	131	8.3	1547	13	US-10-027-632-265136		Sequence 265136, S
11	131	8.3	1547	14	US-10-027-632-265133		Sequence 265133, S
12	131	8.3	1547	14	US-10-027-632-265134		Sequence 265134, S
13	131	8.3	1547	14	US-10-027-632-265135		Sequence 265135, S
14	131	8.3	1547	14	US-10-027-632-265136		Sequence 265136, S
15	107.6	6.8	1254	15	US-10-229-546-3		Sequence 3, Appl

16	107.6	6.8	1622	12	US-10-429-802-21	Sequence 21, Appl
17	107.6	6.8	1622	15	US-10-229-546-1	Sequence 1, Appl
18	107.6	6.8	1633	13	US-10-341-434-187	Sequence 187, App
19	101	6.4	1870	10	US-09-910-059-112	Sequence 112, App
20	101	6.4	2154	10	US-09-910-059-124	Sequence 124, App
21	94.6	6.0	1332	10	US-09-954-456-1141	Sequence 1141, Ap
22	92.8	5.9	1311	14	US-10-200-344-9	Sequence 9, Appl
23	92.8	5.9	1993	12	US-10-274-639-33	Sequence 33, Appl
24	92.8	5.9	2128	14	US-10-200-344-13	Sequence 13, Appl
25	91.6	5.8	936	15	US-10-106-698-1866	Sequence 1866, Ap
26	88.6	5.6	517	15	US-10-198-846-12472	Sequence 12472, A
27	88.6	5.6	1050	14	US-10-200-344-11	Sequence 11, Appl
28	87.6	5.6	1200	15	US-10-200-910-7	Sequence 7, Appl
29	83.6	5.3	945	15	US-10-200-910-3	Sequence 3, Appl
30	83.6	5.3	945	15	US-10-200-910-11	Sequence 11, Appl
31	80.4	5.1	451	15	US-10-060-036-3265	Sequence 3265, Ap
32	80	5.1	1125	9	US-09-888-615-2	Sequence 2, Appl
33	80	5.1	1332	16	US-10-176-306-75	Sequence 75, Appl
34	80	5.1	1603	13	US-10-176-306-73	Sequence 73, Appl
35	80	5.1	1826	16	US-10-252-157-453	Sequence 453, App
36	79.2	5.0	878	15	US-10-198-846-4746	Sequence 4746, Ap
37	78.4	5.0	741	14	US-10-200-344-5	Sequence 5, Appl
38	78.4	5.0	1603	16	US-10-176-306-73	Sequence 73, Appl
39	76.8	4.9	948	9	US-09-888-615-1	Sequence 1, Appl
40	76.8	4.9	1187	15	US-10-198-846-13457	Sequence 13457, A
41	72.8	4.6	1341	9	US-09-925-297-69	Sequence 69, Appl
42	71.6	4.6	1306	9	US-09-923-779-144	Sequence 144, App
43	71.6	4.6	1306	9	US-09-962-436-557	Sequence 557, App
44	70.6	4.5	1254	9	US-09-925-297-58	Sequence 58, Appl
45	70.6	4.5	1260	13	US-10-345-680-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1

```

US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1
US-09-880-107-2396

```

	Query Match	87.2%	Score 1372;	DB 10;	Length 1728;
	Best Local Similarity	90.3%;	Pred. No. 0;		
	Matches 1558;	Conservative	0;	Mismatches	5; Indels 163; Gaps 2;
Qy	1	AGAAATTCGTGTTGGATGAAGCTTTGCAGCCTTGCAGTCCCTGTGTACCATGTTCTCT	60		
Db	3	AGAAATTCGTGTTGGATGAAGCTTTGCAGCCTTGCAGTCCCTGTGTACCATGTTCTCT	62		
Qy	61	TCTGTGAGCAGCATGCTTCGCGTTTCAGAGTGGCCAAAGTTCCTAGCTGCTCTCCCTAGAA	120		
Db	63	TCTGTGAGCAGCATGCTTCGCGTTTCAGAGTGGCCAAAGTTCCTAGCTGCTCTCCCTAGAA	122		

121 CCTCTAGGCAAGTTCAAGTCTCAGAACTCTTACTCAACATATGAGATTGTTCTCTGGC 180
123 CCTCTAGGCAAGTTCAAGTCTCAGAACTCTTACTCAACATATGAGATTGTTCTCTGGC 182
181 AGCGGTAAACAGCTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTTTTGTAAATGCGAT 240
183 AGCGGTAAACAGCTGACCTTATTTGTGAAGAAAAACAAGTCCATTTTTTTGTAAATGCGAT 242
241 CTGATGTGCAATGTGAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 300
243 CTGATGTGCAATGTGAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 302
301 TGGCAGACGTGGAAGATCTTATTTCAACAGCAGATTTTCAACAGCAGATTCAGTCCGCGGAG 360
303 TGGCAGACGTGGAAGATCTTATTTCAACAGCAGATTTTCAACAGCAGATTCAGTCCGCGGAG 362
361 CCTCCGCATCGTACTATGAACAGTATCAGTCACTCAATGAATCTATTTCTTGGATAGAAT 420
363 CCTCCGCATCGTACTATGAACAGTATCAGTCACTCAATGAATCTATTTCTTGGATAGAAT 422
421 TTATACTGAGAGCATCTTGATATGCTTACAAAAATCCACATGATGATCTCATTTTGAGA 480
423 TTATACTGAGAGCATCTTGATATGCTTACAAAAATCCACATGATGATCTCATTTTGAGA 482
481 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAAGCAACAGCCAAAAATGCCATAT 540
483 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAAGCAACAGCCAAAAATGCCATAT 542
541 GGATTTGACTGTGAATCCATGCCAGAGAAATGATCTCTCTGCTTTCTGCTGTGTTCA 600
543 GGATTTGACTGTGAATCCATGCCAGAGAAATGATCTCTCTGCTTTCTGCTGTGTTCA 602
601 TAGGCCAT----- 608
603 TAGGCCATATAACTCAATTTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 662
609 -----A 609
663 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGTTTATGACTACTCATGGAAGAAAGA 722
610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACATCATTTGATCGGACAGACACC 669
723 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACATCATTTGATCGGACAGACACC 782
760 TGAATAGCAACTTTGCTTCCAAACACTGCTGTGAGGAAGGTGCATCCAGTTCTCTCATGCT 729
783 TGAATAGCAACTTTGCTTCCAAACACTGCTGTGAGGAAGGTGCATCCAGTTCTCTCATGCT 842
730 CGGAACCTACTGTGAGCTTTATCCTGAGTCAGAACAGAGTGAAGGCAAGTGGCTAGTT 789
843 CGGAACCTACTGTGAGCTTTATCCTGAGTCAGAACAGAGTGAAGGCAAGTGGCTAGTT 902
790 TCTTGAGAAGAAATATCAACAGATTAAGCATACATCAGCATGCAATTCATCTCCAGC 849
903 TCTTGAGAAGAAATATCAACAGATTAAGCATACATCAGCATGCAATTCATCTCCAGC 962
850 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGCTTC 909
963 ATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGCTTC 1022
910 TAGTACCCAGTGAAGCAAGTTCGCTATTTACAAATAGTAAATACCAAGTATACAC 969
1023 TAGTACCCAGTGAAGCAAGTTCGCTATTTAGAAAACTAGTAAAAATACCAAGTATACAC 1082
970 ATGCCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACGATTTGATCTATG 1029
1083 ATGCCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACGATTTGATCTATG 1142
1030 ATTTGGGCATCAATATTCGT----- 1050
1143 ATTTGGGCATCAATATTCGT----- 1202

1051 -----TTACATCAAAACCCACCTGTAGAGAAGCTTTGGCGCTGTCTCTAAAA 1097
1203 TGGTCCGGAGCGTTACATCAAAACCCACCTGTAGAGAAGCTTTGGCGCTGTCTCTAAAA 1262
1098 TAGCTTGGCATGTCAATAGGAATGTTTAAATGATGCCCTGATTTTATCATCTTGTCTCCGTAT 1157
1263 TAGCTTGGCATGTCAATAGGAATGTTTAAATGATGCCCTGATTTTATCATCTTGTCTCCGTAT 1322
1158 TTTAATTTACTGATTTCCAGCAAGCAAAATCATTTGATCAGATTTATTTTAAAGTTTATC 1217
1323 TTTAATTTACTGATTTCCAGCAAGCAAAATCATTTGATCAGATTTATTTTAAAGTTTATC 1382
1218 CGTAGTTTTGATAAAAGATTTTCTATTTCTTGGTTCTGTCAAGAACCTTAATAAGTGCT 1277
1383 CGTAGTTTTGATAAAAGATTTTCTATTTCTTGGTTCTGTCAAGAACCTTAATAAGTGCT 1442
1278 ACTTTCCCATTAAGGCAGACTAGGGTTCAATGCTTTTATCCCTTTAAAAAAAATTTGTAA 1337
1443 ACTTTGCCATTAAGGCAGACTAGGGTTCAATGCTTTTATCCCTTTAAAAAAAATTTGTAA 1502
1338 AAGCTAGTACTACTATTTTCTTTCATTTTTCGACGTTTTCGACCTAGCATCTCAAGCACT 1397
1503 AAGCTAGTACTACTATTTTCTTTCATTTTTCGACGTTTTCGACCTAGCATCTCAAGCACT 1562
1398 TTCGACGTTTTCGACTAGCCATCTCAAGCAAGTTTTAAATCAAGATCATCTCAAGCTGATCAT 1457
1563 TTCGACGTTTTCGACTAGCCATCTCAAGCAAGTTTTAAATCAAGATCATCTCAAGCTGATCAT 1522
1458 TGGATCCTACTCAACAAAAAGGAGGTGTGAGAGTACATTAAGAGATTTTCCTCTCCAA 1517
1623 TGGATCCTACTCAACAAAAAGGAGGTGTGAGAGTACATTAAGAGATTTTCCTCTCCAA 1682
1518 TTTTCAATTAATTTCTTCTCTCTTTTAAAAAAAATTTTAAAA 1563
1683 TTTTCAATTAATTTCTTCTCTCTTTTAAAAAAAATTTTAAAA 1728

RESULT 2

US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAPI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Papio hamadryas
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1269)
US-10-379-836-1

Query Match 54.6%; Score 858.8; DB 13; Length 1272;
Best Local Similarity 83.9%; Pred. No. 5.8e-208;
Matches 1067; Conservative 0; Mismatches 42; Indels 163; Gaps 2;

18 ATGAAGCTTTGACGCTTGCAGCTTCAGTTCCTGTACCCATTTGTTCTCTTCTGTGAGCAGCATGTC 77
1 ATGAAGCTTTGACGCTTGCAGCTTCCTGTGAGCTTCCTGTGAGCAGCATGTC 60
78 TTCGCGTTCCAGAGTGGCCAAAGTTCTAGTGTCTTCTTCCAGAACCTCTTAGGCAAGTTCAA 137
61 TTCGCGTTCCAGAGTGGCCAAAGTTCTAGTGTCTTCTTCCAGAACCTCTTAGGCAAGTTCAA 120

Qy	138	GTTCCTACAGAACTCTTACATCAACAATATGAGATTGTTCTCTGGCAGCCGGTAACACGCTGAC	197
Db	121	GTGCTTACAGAACTCTTACTTCAACAATATGAGATTGTTCTCTGGCAGCCGGTAACACGCGAC	180
Qy	198	CTTATTGTGGAAGAAAACAAGTCCATTTTTTTGTAAATGATCTGATGTGCAAAATGTG	257
Db	181	CTTATTGTGAAGAAAACAAGTCCATTTTTTTGTAAATCATCTGATGTGCAAAATGTG	240
Qy	258	AAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTCTTCTGCTGGCAGAGCGTGAAGAT	317
Db	241	AAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTCTTCTGCTGGCAGATGTGGAAGAT	300
Qy	318	CTTATTCAAAGCAGAGATTTCCAAAGACACAGTCAGCCCGCCGAGCCTCCGCATCGTACTAT	377
Db	301	CTTATTCAAAGCAGAGATTTCCAAAGACACAGTCAGCCCGCCGAGCCTCCGCATCGTACTAT	360
Qy	378	GAAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATAACTGAGAGGCAT	437
Db	361	GAAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATAACTGAGAGAT	422
Qy	438	CCTGATATGCTTACAAAATCCACATTTGGATCCTCATTTGAGAAAGTACCCACTCTATGTT	497
Db	421	CCTGATATGCTTACAAAATCCACATTTGGATCCTCTATGAGAGCACCACTTTATGTT	480
Qy	498	TTAAAGGTTTCTGGAAGAAGAACAAACAGCCAAAATAAGCCATATGGAATGACTGTGGAATC	557
Db	481	TTAAAGGTTTCTGGAAGAAGAACAAACAGCCAAAATAAGCCATATGGAATGACTGTGGAATC	540
Qy	558	CATGCCAGAGAAATGGATCTCTCTGCTTTCTGCTTTGTTGTTTCATAGGCCAT-----	608
Db	541	CATGCCAGAGAAATGGATCTCTCTGCTTTCTGCTTTGTTGTTTCATAGGCCATATAAAGTAA	600
Qy	609	-----	608
Db	601	TACTACGGGATAATAGGGGAATATACCAATCTTCTGAGGCATGTGGATTTCTATGTTATG	660
Qy	609	-----	626
Db	661	CCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAGAAATCGAATGTGAGAGAAAG	720
Qy	627	AACCGTTCTTCTATGCGAACAAATCATGATCGGAACAGACCTGGAATAGCAACTTTGTC	686
Db	721	AACCGTTCTTCTATGCGAACAAATCGTTGCAATCGGAACAGACCTGGAACAGGAACCTTGG	780
Qy	687	TCCAAAACACTGGTGTGAGGAAGTGCATCCAGTTTCTCATGCTCGGAAACCTCTGTGGGA	746
Db	781	TCCAAAACACTGGTGTGAGGAAGTGCATCCAGTTTCTCATGCTCGGAAACCTCTGTGGGA	840
Qy	747	CTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTTGAGAGAAATATC	806
Db	841	CTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTTGAGAGAAATATC	900
Qy	807	AACAGATTTAAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTTCATAT	866
Db	901	AACACATTTAAAGCATACATCAGCATGCATTCATCTCCAGCATATCGTGTTCATAT	960
Qy	867	TCCTATACAGAACTAAAGCAAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA	926
Db	961	TCCTATACTCGAACAAAGCAAAAGACCAAGAGAAATGTCTCTAGTAGCCAGTGAAGCA	1020
Qy	927	GTTCGTGCTATTGACAAACTAGTAAATAACAGGTATACATGATGCCATGGCTCAGAA	986
Db	1021	GTTCGTGCTATTGACAAACCCAGTAAATAATCAGGTATACATGATGCCGTGGCTCAGAA	1080
Qy	987	ACCTTATACCTAGTCTCTGAGGTGGGGAAGATTGGATCTATGATTTGGGGCATCAAAATAT	1046
Db	1081	ACCTTATACCTAGTCTCTGAGGTGGGGAAGATTGGATCTATGATTTGGGGCATCAAAATAT	1140
Qy	1047	TCGT-----	1054
Db	1141	TCGTTTACAATTTGAATCTCGGAGATACGGGCAAAATACGGAATTTCTGCTGCGCTGACGGTTAC	1200
Qy	1055	ATCAAAACCCACCTGTAGAGAGCTTTTGGCCGCTGTCTCTAAAATAGCTTTGGCATGTCAATT	1114

[illegible]

Query Match	54.3%	Score 853.8	DB 9	Length 1400	
Best Local Similarity	93.3%	Pred. No. 1.2e-206			
Matches	934	Conservative 1	Mismatches 13	Indels 53	Gaps 2
Qy	607	ATAATCGAATGTCGGAGAAAGAACCGTTCTTCTTATCGGAACAATCATTCGATCGGAACAG	666		
Db	177	AGAATCGAATGTCGGAGAAAGAACCGTTCTTCTTATCGGAACAATCATTCGATCGGAACAG	236		
Qy	667	ACCTGAATACCAACTTGTCTCCAAACACCTGGTGTGAGGAAGGTGCATCCAGTTCTCTCAT	726		
Db	237	ACCTGAATACCAACTTGTCTCCAAACACCTGGTGTGAGGAAGGTGCATCCAGTTCTCTCAT	296		
Qy	727	GCTCGGAAACCTACTGTGGACTTTTATCCTGAGTCAGAACGAGAAAGTGAAGGCACTGGCTA	786		
Db	297	GCTCGGAAACCTACTGTGGACTTTTATCCTGAGTCAGAACGAGAAAGTGAAGGCACTGGCTA	356		
Qy	787	GTTCCTTGAGAAAGAAATATCAACAGATTAAAGCATACATCAGCATGCATTCACTCTCC	846		
Db	357	GTTCCTTGAGAAAGAAATATCAACAGATTAAAGCATACATCAGCATGCATTCACTCTCC	416		
Qy	847	AGCATATAGTGTTCATATTCCTATACAGAACTAAAGCAAAAGCACTGAGGAACCTGT	906		
Db	417	AGCATATAGTGTTCATATTCCTATACAGAACTAAAGCAAAAGCACTGAGGAACCTGT	476		
Qy	907	CTCTAGTAGCCAGTGAAGCAGTTCTGTCATTGACAAAACTAGTAAAAAATACCAAGTATA	966		
Db	477	CTCTAGTAGCCAGTGAAGCAGTTCTGTCATTGAGAAAACTAGTAAAAAATACCAAGTATA	536		
Qy	967	CACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTCGAGGTGGGACGATTGGATCT	1026		
Db	537	CACATGGCCATGGCTCAGAAACCTTATACTAGTCTCTCGAGGTGGGACGATTGGATCT	596		
Qy	1027	ATGATTTGGGCATCAAAATATTCGT	1050		
Db	597	ATGATTTGGGCATCAAAATATTCGTTCACAACTTCGAGNATACGGGCACATACGGAT	656		
Qy	1051	-----TTACATCAAAACCACTGTAGAAAGCTTTTGGCGTGCTCTTA	1094		
Db	657	TCCTGTGCGGAGCGTTACATCAAAACCACTGTAGAAAGCTTTTGGCGTGCTCTTA	716		

QY 1095 AATAGCTTGGCATGTCATTAGGAATGTTAAATGCCCTCGATTTTATCATTTCTGCTTCGG 1154
DB 717 AATAGCTTGGCATGTCATTAGGAATGTTAAATGCCCTCGATTTTATCATTTCTGCTTCGG 776
QY 1155 TATTTAAATTTACTGATTCAGCAAGACCAATCAATGTATCATGATTTATTTAAAGTTTT 1214
DB 777 TATTTAAATTTACTGATTCAGCAAGACCAATCAATGTATCATGATTTATTTAAAGTTTT 836
QY 1215 ATCGTAGTTTGGATTAAGAAATTTCTTATTCCTTGGTTCTGTCTGTCAGAGAACCTTAATAGT 1274
DB 837 ATCGTAGTTTGGATTAAGAAATTTCTTATTCCTTGGTTCTGTCTGTCAGAGAACCTTAATAGT 896
QY 1275 GCTACTTTGGCATTTAAGCGACACTAGGTTTCATGCTCTTTTACCCTTTTAAAGAAATTTG 1334
DB 897 GCTACTTTGGCATTTAAGCGACACTAGGTTTCATGCTCTTTTACCCTTTTAAAGAAATTTG 955
QY 1335 TAAAGTCTAGTTACCTACTTTTTTTTGTGATTTTTCGACGTTTGACTAGCCATCTCAAGCA 1394
DB 956 TAAAGTCTAGTTACCTACTTTTTTTTGTGATTTTTCGACGTTTGACTAGCCATCTCAAGCA 1015
QY 1395 ACTTTCAGCTTTGACTAGGATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGAT 1454
DB 1016 ACTTTCAGCTTTGACTAGGATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGAT 1075
QY 1455 CATTTGGATCTACTCAACAAAGAGGTTGTCAGAGTACATTTAAAGATTTTCTGCTCC 1514
DB 1076 CATTTGGATCTACTCAACAAAGAGGTTGTCAGAGTACATTTAAAGATTTTCTGCTCC 1135
QY 1515 AATTTTCAATAAATTTCTTCTCTCTCTTTTAAAAAATAA 1555
DB 1136 AATTTTCAATAAATTTCTTCTCTCTCTTTAGAAATACAA 1176

RESULT 4

US-09-960-352-14595
; Sequence 14595, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagabhan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14595
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10
US-09-960-352-14595

Query Match 21.1%; Score 331.2; DB 10; Length 416;
Best Local Similarity 87.3%; Pred. No. 6e-74;
Matches 363; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 167 GATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGAAACCAAGTCCATTT 226
DB 1 GATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGAAACCAAGTCCATTT 60
QY 227 TTTTGTAAATGCACTGTCGATGTCGACCAATGTGAAGCCCAATTTAAATGTGAGCGGAATTC 286
DB 61 TTTTGTGAATGCACTGTCGATGTCGACCAATGTGAAGCCCAATTTAAATGTGAGCGGAATTC 120
QY 287 ATGCAAGTCTTCTGTCGACAGCTGGAAGATCTTTTCAACAGCAGATTTTCAACAGCAC 346
DB 121 ATTCAAGGTCCTGGTGAAGAAATGTGAAGATCTTTATCCGGCAGCAGACTTTCCAATGACAC 180
QY 347 AGTCAGGCCCCGAGCTCCCGATCGTACTATGAACAGTATCATCTCACTAAATGAATCTTA 406

DB 181 CATCAGCCCCGGGCATCTCTCTCTACTATGAACAGTATCACTCACTAAATGAGATCTA 240
QY 407 TTCTTGGATAGAAATTTATTAACGTAGAGGATCTCTGATATGCTTTACAAAATCCACATGG 466
DB 241 TTCTTGGATAGAAATTTATGACTGAGCGGTATCTCTGATATGTTTGAAGAAATCCACATGG 300
QY 467 ATCCTCATTTGAGAACTACCCACTCTATGTTTAAAGGTTTCTGGAAAAGAACAAACAGC 526
DB 301 ATCCTCATGAGAACTACCCACTTTATGTTTAAAGGTTTCTAAAGAAAGAACAAAGGC 360
QY 527 CAAAAATGCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
DB 361 CAAAAATGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 416

RESULT 5

US-09-960-352-8425
; Sequence 8425, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagabhan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8425
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 36-LIB34-005-Q1-E1-A12
US-09-960-352-8425

Query Match 18.2%; Score 286.8; DB 10; Length 397;
Best Local Similarity 83.9%; Pred. No. 1.2e-62;
Matches 324; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 12 GTTGGGATGAAGCTTTTGACGCTTTCAGTCTTGTACCCATTTCTTCTCTGAGCAG 71
DB 12 GTGCGGATGAAGCTTTTATAGCTTGGAGTCTTGTGCGCACCGTCTCTTCTGTTGGGAG 71
QY 72 CATGCTTTCGCGTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAA 131
DB 72 CATGCTTTCGCGTTTCAGAGTGGCCAAAGTTTATCTGCTCTTCTAGAACCTCTAGGCAA 131
QY 132 GTTCAAAGTCTACAGATCTTACTACACATATGAGATTTCTTCTGGCAGCGCGTAACA 191
DB 132 GTTCAAAGTCTACAGATTTTACTACACATATGAGATTTCTTCTGGCAGCGCGTAGCA 191
QY 192 GCTGACCTTATGTTGAAGAAACCAAGTCCATTTTGTAAATGCATCTGATGTCGAC 251
DB 192 GCTGAATATATGTTGAAGGATACGAAGTCCATTTTGTGAATGATCTGATGAAGC 251
QY 252 AATGTGAAGCCCATTTAAATGTGAGCGGAATTTCCATGCAAGTCTTCTGTTGGCAGCGTG 311
DB 252 AATGTGAAGCCCATTTAAATGTGAGCGGAATTTCCATGCAAGTCTTCTGTTGGAAATGTG 311
QY 312 GAAGATCTTATTTCAACAGCAGATTTTCAACGACACAGTACAGCCCCCGAGCTCCGATCG 371
DB 312 GAAGATCTTATTCGGCAGCAGACTTCCAATGACACCATCAGCCCCCGGCGCATCTCTCTCC 371
QY 372 TACTATGAACAGTATCACTCACTAAA 397
DB 372 TACTATGAACAGTATCACTCACTAAA 397

RESULT 6

US-09-917-800A-468/c

```
; Sequence 468, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 468
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(431)
; OTHER INFORMATION: n = a or c or g or t
; US-09-917-800A-468

Query Match      8.7%; Score 137.2; DB 10; Length 431;
Best Local Similarity 67.5%; Pred. No. 1.4e-24;
Matches 268; Conservative 0; Mismatches 69; Indels 60; Gaps 3;

QY      894 CATGAGAACTGCTCTAGTAGCAGTGAAGCAGTTCGTCTATTTGACAAACTAGTAA 953
Db      431 CATGAGAACTGCTCTAGTAGCAGTGAAGCAGTTCGTCTATTTGACAAACTAGTAA 372
QY      954 AATACCAAGTATACACATGCGCATGCTCAGAACTTATACCTAGTCTCTGAGGTGG 1013
Db      371 AACACCAAGTATACACATGCGCATGCTCAGAACTTATACCTAGTCTCTGAGGTTC 312
QY      1014 GACGATTTGATCTATGATTTGGGATCAAAATTTTGGTTTA----- 1053
Db      311 GATGATTTGATCTATGATTTGGGATCAAAATTTTGGTTTACGATTTGAACTTCGGGATACA 252
QY      1054 -----CATCAAAACCCACCTGTAGAGAGCTTTT 1081
Db      251 GGCAGATACGGGTTCTTGCTGCTGAGAGATTCATCAAAACCCACTTGGCGAGAAGCTTGG 192
QY      1082 GCGCTGCTCTAAATAGTGTGCGATGCTATTAGGAATGTTTAATGCCCTGATTTAT 1141
Db      191 GCGCGAGTCTCTAAATAGTGTGCGATGCTATTAGGAATGTTTAATGCCCTGATTTAT 138
QY      1142 CATTCGCTTCCTGATTTTAAATTTACTGATTCAGCAAGCAAAATCATTTGATCATGAT- 1200
Db      137 CTCTGCTCTCAATACATTTTATTTATTTGATTTGAGCAACACTAAATTTGTTGCACTAGCTT 78
QY      1201 -TATTTTAAAGTTTATCCCTAGTTTGTGATAAAAGAT 1236

Db      77 CTAAGTTTAAATCAGTTTCTCTGTTGTTTGTGAAGAAT 41

RESULT 7
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265133
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-265133

Query Match      8.3%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      702 GAGGAAGTGCATCCAGTTCCTCATCTCGGAACCTACTGTGGACTTTATCTCTGATCA 761
Db      1417 GAGGAAGTGCATCCAGTTCCTCATCTCGGAACCTACTGTGGACTTTATCTCTGATCA 1476
QY      762 GAACCAAGTGAAGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 821
Db      1477 GAACCAAGTGAAGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 1536
QY      822 TACATCAGCAT 832
Db      1537 TACATCAGCAT 1547

RESULT 8
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
```



```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265134
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265134

Query Match      8.3%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 761
Db 1417 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 1476

QY 762 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 821
Db 1477 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 822 TACATCAGCAT 832
Db 1537 TACATCAGCAT 1547

RESULT 9
US-10-027-632-265135
; Sequence 265135, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265136
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

Query Match      8.3%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 761
Db 1417 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 1476

QY 762 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 821
Db 1477 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 822 TACATCAGCAT 832
Db 1537 TACATCAGCAT 1547
```

```
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265135

Query Match      8.3%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 761
Db 1417 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 1476

QY 762 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 821
Db 1477 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 822 TACATCAGCAT 832
Db 1537 TACATCAGCAT 1547

RESULT 10
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265136
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

Query Match      8.3%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 761
Db 1417 GAGGAAGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGGACTTTATCTCTGAGTCA 1476

QY 762 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 821
Db 1477 GAACCAAGAGTGAAGCAGTGGCTAGTTCTTTCGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 822 TACATCAGCAT 832
Db 1537 TACATCAGCAT 1547
```

RESULT 11

```
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265133
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265133
```

```
Query Match      8.3%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAGGAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCTCTGAGTCA 761
    |||||||
Db 1417 GAGGAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCTCTGAGTCA 1476

QY 762 GAACCAAGTGAAGGCGAGTGGCTAGTTCTTCTGAGAAGAAATATCAACAGATTAAAGCA 821
    |||||||
Db 1477 GAACCAAGTGAAGGCGAGTGGCTAGTTCTTCTGAGAAGAAATATCAACAGATTAAAGCA 1536

QY 822 TACATCAGCAT 832
    |||||||
Db 1537 TACATCAGCAT 1547
```

RESULT 12

```
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265134
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265134
```

```
Query Match      8.3%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265134
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265134

Query Match      8.3%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 GAGGAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCTCTGAGTCA 761
    |||||||
Db 1417 GAGGAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCTCTGAGTCA 1476

QY 762 GAACCAAGTGAAGGCGAGTGGCTAGTTCTTCTGAGAAGAAATATCAACAGATTAAAGCA 821
    |||||||
Db 1477 GAACCAAGTGAAGGCGAGTGGCTAGTTCTTCTGAGAAGAAATATCAACAGATTAAAGCA 1536

QY 822 TACATCAGCAT 832
    |||||||
Db 1537 TACATCAGCAT 1547
```

RESULT 13

```
US-10-027-632-265135
; Sequence 265135, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265135
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265135
```

```
Query Match      8.3%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 1.1e-22;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	702	GAGGAAGTGCATCCAGATTCTCATGCTCGAAA	CCTACTGTGAGATTTTATCCTGAGTCA	761
Db	1417	GAGGAAGTGCATCCAGATTCTCATGCTCGAAA	CCTACTGTGAGATTTTATCCTGAGTCA	1476
Qy	762	GAACCGAAGTGAAGGCAGTGGCTAGTTCCTTGAGAA	GAATATCAACCAAGATTAAGCA	821
Db	1477	GAACCGAAGTGAAGGCAGTGGCTAGTTCCTTGAGAA	GAATATCAACCAAGATTAAGCA	1536
Qy	822	TACATCAGCAT	832	
Db	1537	TACATCAGCAT	1547	

RESULT 14

```

RESULT 14
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265136
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

```

[illegible]

RESULT 15

RESULTS 15
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 10:44:24 ; Search time 3636.93 Seconds
(without alignments)
10511.864 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattctgttgatg.....aaaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY NUC

.Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	717.8	45.6	973	9	AL536068
2	690.8	43.9	932	13	BX451532
3	663.2	42.2	971	9	AL536069
4	656	41.7	1446	11	AK004045

5	622.8	39.6	780	10	BG616950
6	615.6	39.1	806	10	BG567264
7	604.2	38.4	788	10	BG618629
8	594.8	37.8	702	2	BSM089979
9	581.2	36.9	747	10	BG566051
10	579.2	36.8	725	10	BG618239
11	561.8	35.7	738	10	BG616456
12	555.2	35.3	735	9	AV646979
13	536.2	34.1	541	14	CB157619
14	532.2	33.8	750	10	BG618813
15	523.8	33.3	691	10	BG568240
16	510.4	32.4	623	9	AV694058
17	509.4	32.4	583	9	AV693037
18	509.4	32.4	602	9	AV692032
19	507.4	32.3	621	9	AV698425
20	501.2	31.9	601	9	AV658390
21	491.6	31.3	643	9	AV688323
22	483.4	30.7	485	13	BX091627
C 23	466	29.8	589	9	AV719484
24	466	29.6	558	9	AV662298
C 25	465.4	29.6	611	14	N63924
26	465	29.6	603	10	AW950700
C 27	464	29.5	586	9	AV719308
28	461.8	29.4	688	9	AV651709
C 29	460.2	29.3	495	9	AI271641
C 30	460.2	29.3	585	9	AV720894
C 31	460.2	29.3	611	9	AA700377
C 32	457.8	29.1	537	14	N98450
33	455	28.9	577	14	CB154355
34	450.4	28.6	462	9	AV681868
35	444.2	28.2	563	14	CB162759
C 36	442	28.1	576	9	AV720000
C 37	441.6	28.1	623	9	AV658178
C 38	438.6	27.9	586	9	AV719997
39	436	27.7	510	14	CB156851
40	432.8	27.5	1004	14	BY704293
41	427.4	27.2	455	14	W88434
C 42	422.4	26.9	593	14	W89198
C 43	420.6	26.7	747	10	BG565100
C 44	418.4	26.6	450	9	AI076434
45	417.4	26.5	958	9	AI322676

ALIGNMENTS

RESULT 1.
AL536068/c
LOCUS AL536068 Homo sapiens FETAL BRAIN Homo sapiens CDNA clone
DEFINITION CS0DF022YE21 3-PRIME, mRNA sequence.
ACCESSION AL536068.2 GI:31260938
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 973)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12799561.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6944.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF022AC11NP1&cluster=6944.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DF022AC11NP1.

FEATURES
source
Location/Qualifiers
i. 973
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 284 a 193 c 188 g 250 t 58 others
ORIGIN

Query Match 45.6%; Score 717.8; DB 9; Length 973;
Best Local Similarity 86.0%; Pred. No. 2.5e-77;
Matches 793; Conservative 44; Mismatches 31; Indels 54; Gaps 3;

607 ATATCGAATGTGGAGAAACCGTTCTTTCTATGCGAACAAATCATTCGATCGGAACAG 666
608 AGAATCGAATGTGGAGAAACCGTTCTTTCTATGCGAACAAATCATTCGATTCGGAACAG 897
609 ACCTGAATAGCAATTTGTCTCCAAACACCTGGTGTGAGGAAGTGATCCAGTTCTTCAT 726
610 ACCTGAATAGCAATTTGTCTCCAAACACCTGGTGTGAGGAAGTGATCCAGTTCTTCAT 837
611 GCTCGGAACCTACTGTGGACTTTATCTGAGTCAGAACAGAGAGTGAAGCGATGGCTA 786
612 GCTCGGAACCTACTGTGGACTTTATCTGAGTCAGAACAGAGAGTGAAGCGATGGCTA 777
613 GTTCTCTGAGAGAAATATCAAA-CCAGATTAAAGCATACATCAGCATGATTCATATCTCC 845
614 GTTCTCTGAGAGAAATATCAAGCCAGATTAAAGCATACATCAGCATGATTCATATCTCC 717
615 CAGCATATAGTGTTCATATTCCTATACAGAAATGAAAGCAAGCATGAGGAAGTGG 905
616 CAGCATATAGTGTTCATATTCCTATACAGAAATGAAAGCAAGCATGAGGAAGTGG 657
617 TCTCTAGTACCGAGTGAAGCAGTTCGCTGCTATTGACAAACTAGTAAATACAGGTAT 965
618 TCTCTAGTACCGAGTGAAGCAGTTCGCTGCTATTGACAAACTAGTAAATACAGGTAT 597
619 ACATATGCGCATGGCTCAGAAACCTTATACCTAGTCTCTGAGTGGGAGCATGATGATC 1025
620 ACATATGCGCATGGCTCAGAAACCTTATACCTAGTCTCTGAGTGGGAGCATGATGATC 537
621 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCACATACGGA 1055
622 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCACATACGGA 477
623 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 1093
624 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 417
625 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 1153
626 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 357
627 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 1213
628 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 297
629 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 1273
630 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 237
631 TATGATTGGGCATCAAAATATTCGTTTACAAATGAACTTCGAGTTACGGGCATCTCT 1333

236 TGCTWTWTTGCCATTAAAGCGWAGCTWGGGTTCATGTATTTTACCCCTTT-WAAAAAAATT 178
1334 GTAAAGTCTAGTACTACTACTTTCTTTCATTTTCGAGTTTCGAGTTCAGTTCAGTCAAGC 1393
177 KTWAAAGTCTWGTTCCTACTACTTTCTTTCATTTTCGAGTTTCGAGTTCAGTTCAGTCAAGC 118
1394 AACTTTTCGAGTTTCGAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 1453
117 SGSKTSGGGCTTTCGAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 58
1454 TCATTCGATCTCTACTCAACAAA 1475
57 TCATTCGATCTCTANTNNAACA 36

RESULT 2
BX451532 932 bp mRNA linear EST 22-MAY-2003
LOCUS
DEFINITION BX451532 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF022YE21 5-PRIME, mRNA sequence.
ACCESSION BX451532
VERSION BX451532.1 GI:31024255
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 932)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6944.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAF0162B11_AF01523_1&cluster=6944.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0BAF0162B11_AF01523_1.

FEATURES
source
Location/Qualifiers
i. 932
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT 281 a 189 c 170 g 283 t 9 others
ORIGIN

Query Match 43.9%; Score 690.8; DB 13; Length 932;
Best Local Similarity 91.6%; Pred. No. 4.5e-74;
Matches 785; Conservative 0; Mismatches 18; Indels 54; Gaps 3;

720 TCCTCATGCTCGAAACCTACTGTGACCTTTATCTGAGTCAGAACCAAGTGAAGGCA 779
14 TCCTCATGCTCGAAACCTACTGTGACCTTTATCTGAGTCAGAACCAAGTGAAGGCA 73
780 GTGGCTAGTTTCTTGAGAGAAATATCAACAGATTAAAGCATACATCAGCATGATTC 839
74 GTGGCTAGTTTCTTGAGAGAAATATCAACAGATTAAAGCATACATCAGCATGATTC 133
840 TACTCCAGCATATAGTGTGTTTCCATATTCCTATACAGAAAGTAAAGCAAGACCATGAG 899

```
Db 134 TACTCCAGATATAGTGTTCATATATCTTATACAGAGTAAAGCAAGACCATGAG 193
QY 900 GAACTGTCTCTAGTACCCAGTGAAGCAGTTCGTGCTATTGACAAAACCTAGTAAATATACC 959
Db 194 GAACTGTCTCTAGTACCCAGTGAAGCAGTTCGTGCTATTGAGAAAACCTAGTAAATATACC 253
QY 960 AGTATATACATGCGCATGCGTCCAGAAACCTTATACCTAGTCTCTGAGAGTGGGAGAT 1019
Db 254 AGTATATACATGCGCATGCGTCCAGAAACCTTATACCTAGTCTCTGAGAGTGGGAGAT 313
QY 1020 TGATCTATGATTTGGGCAATCAATATTCGT----- 1050
Db 314 TGGATCTATGATTTGGGCAATCAATATTCGTTTCAATTTGAATCTCGAGATCGGGACA 373
QY 1051 -----TTATCATCAACCCACCTGTAGAGAAGCTTTTGGCGCT 1087
Db 374 TACGGATTCCTGTCGCGGAGCGTTACATCAACCCACCTGTAGAGAAGCTTTTGGCGCT 433
QY 1088 GTCTCTAAATAGCTTGGCATGTCATAGGAATGTTTAAATGCCCCCTGATTTATCATTTCT 1147
Db 434 GTCTCTAAATAGCTTGGCATGTCATAGGAATGTTTAAATGCCCCCTGATTTATCATTTCT 493
QY 1148 GCTTCGCTATTTTAAATTTACTGATTCAGCAAGCAACCAATCATTTGATCAGATTTATTTT 1207
Db 494 GCTTCGCTATTTTAAATTTACTGATTCAGCAAGCAACCAATCATTTGATCAGATTTATTTT 553
QY 1208 AAGTTTATCCGTAAGTTTGTATAAAGATTTTCTTATTCCTGTTCTGTGAGAGAACCT 1267
Db 554 AAGTTTATCCGTAAGTTTGTATAAAGATTTTCTTATTCCTGTTCTGTGAGAGAACCT 613
QY 1268 AATAAGTCTACTTTGCCATTAAGGAGACTAGGGTTCATGCTTTTACCCTTTTAAAAA 1327
Db 614 AATAAGTCTACTTTGCCATTAAGGAGACTAGGGTTCATGCTTTTACCCTTTTAAAAA 673
QY 1328 AATAAGTCTACTTTGCCATTAAGGAGACTAGGGTTCATGCTTTTACCCTTTTAAAAA 1387
Db 674 AATAAGTCTACTTTGCCATTAAGGAGACTAGGGTTCATGCTTTTACCCTTTTAAAAA 732
QY 1388 TCAAGCAACTTTGCGAGTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCA 1447
Db 733 TCAAGCAACTTTGCGAGTTTGACTAGCCATCTCAAGCAAGTTTAAATCAAGATCATCTCA 792
QY 1448 CGCTGATCATTTGATCCTACTCAACAAAGAGAGGTTGTCAGAGTACATTAAGATTT 1507
Db 793 CGCTGATCATTTGATCCTACTCAAC-AAAGAGAGGTTGTCAGAGTACATTAAGATTT 851
QY 1508 CTGCTCCAAATTTTCAA 1524
Db 852 CTGCTCCAAATTTTCAA 868

RESULT 3
AL536069 971 bp mRNA linear EST 31-MAY-2003
LOCUS AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0DF022YE21 5-PRIME, mRNA sequence.
ACCESSION AL536069
VERSION AL536069.2 GI:31260939
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 971)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.:
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12799562.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
```

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6944.r. For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DF022AC11Q1&cluster=6944.r>. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF022AC11Q1P1.

FEATURES

source

1...971
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pcMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime and enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pcMVSPORT 6 vector. Library was not normalized."

BASE COUNT 282 a 201 c 211 g 272 t 5 others
ORIGIN

Query Match 42.2%; Score 663.2; DB 9; Length 971;
Best Local Similarity 86.8%; Pred. No. 9e-71;
Matches 786; Conservative 3; Mismatches 6; Indels 111; Gaps 1;

QY 15 GGGATGAGCTTTGCGAGCTTCGAGTCCCTGCTGCTTGTACCAATGTTCTCTCTGTGAGCAGCAT 74
Db 62 GGGAGAGCTTTGCGAGCTTCGAGTCCCTGCTGCTTGTACCAATGTTCTCTCTGTGAGCAGCAT 121
QY 75 GTCTTCGCTTCGAGAGTGGCCAAAGTTCTAGTCTCTTCTAGAGAACCTCTAGGCAAGTT 134
Db 122 GTCTTCGCTTCGAGAGTGGCCAAAGTTCTAGTCTCTTCTAGAGAACCTCTAGGCAAGTT 181
QY 135 CAAGTCTTACAGATCTTACTACACATATGAGATGTTCTCTGCGAGCGGTAAACAGCT 194
Db 182 CAAGTCTTACAGATCTTACTACACATATGAGATGTTCTCTGCGAGCGGTAAACAGCT 241
QY 195 GACCTTATTTGAGAGAAAACCAAGTCCATTTTTTGTAAATGCATCTGATGTCGCAAT 254
Db 242 GACCTTATTTGAGAGAAAACCAAGTCCATTTTTTGTAAATGCATCTGATGTCGCAAT 301
QY 255 GTGAAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGAGAGTGGAA 314
Db 302 GTGAAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGAGAGTGGAA 361
QY 315 GATCTTATTTCAACAGCAGATTTTCCAAACGACAGTCAGCCCCCGAGCTCCCGCATCGTAC 374
Db 362 GATCTTATTTCAACAGCAGATTTTCCAAACGACAGTCAGCCCCCGAGCTCCCGCATCGTAC 421
QY 375 TATGAACAGTATCAGTCACTAAATGAATCTATTCTTGGATAGATTAATTAACGTGAGG 434
Db 422 TATGAACAGTATCAGTCACTAAATGAATCTATTCTTGGATAGATTAATTAACGTGAGG 481
QY 435 CATCTGTATGCTTACAAAATTCACATTTGGATCTCTCATTTTGAGAGTACCCACTCTAT 494
Db 482 CATCTGTATGCTTACAAAATTCACATTTGGATCTCTCATTTTGAGAGTACCCACTCTAT 541
QY 495 GTTTTAAAGTTTCTGGAAGAAACAAACAGCCAAAATATGCCATATGGATTTGATCTGGA 554
Db 542 GTTTTAAAGTTTCTGGAAGAAACAAACAGCCAAAATATGCCATATGGATTTGATCTGGA 601
QY 555 ATCCATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGTTGTTTATAGGCCAT----- 608
Db 602 ATCCATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGTTTATAGGCCATATAACT 661
QY 609 ----- 608
Db 662 CAATCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTT 721

```

609 -----AATCGAATGTGGAGA 623
Db      722 ATGCGGTGTTAATGTGGATGTTATGACTACTCATCGGAAGAAGATCAATGTGGAGA 781
Qy      624 AGAAGCGTTCTTCTATGCGAACAATCATTTGATCGAAGACAGACCTGAATAGCAACTTT 683
Db      782 AGAAGCGTTCTTCTATGCGAACAATCATTTGATCGAAGACAGACCTGAATAGCAACTTT 841
Qy      684 GTCTCCAAACACTGGTGTAGGAGGTGCATCCATGCTCTGCGAAACCTACTGT 743
Db      842 GCTTCCAAACACTGGTGTAGGAGGTGCATCCATGCTCTGCGAAACCTACTGT 901
Qy      744 GGACTTTATCTGAGTCAGAACAGAGTGAAGCGAGTCTAGTTCTTCTGAGAAGAAAT 803
Db      902 GGACTTTATCTGAGTCAGAACAGAGTGAAGCGAGTCTAGTTCTTCTGAGAAGAAW 961
Qy      804 ATCAAC 809
Db      962 ATWAAC 967

RESULT 4
AK004045
LOCUS
DEFINITION
AK004045 1446 bp mRNA linear HTC 05-DEC-2002
Mus musculus 18-day embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110032P04 product:carboxypeptidase B2
(plasma), full insert sequence.
ACCESSION
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kiteunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE
4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadoya,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Peeble,G., Schiavone,T.,
Quackenbush,J., Schriml,L.M., Staabli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,

```

```

Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

REFERENCE
5 The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1446)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hirooka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
15' GAGAGAGAGCGCCGACCTCGAGTTTCTTTTNN 3', cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence[5'
GAGAGAGAGAGTCCCAAGCTCAATTAATTAATTAACCCCCCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.
FEATURES
Location/Qualifiers
1..1446
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM.DB:1110032P04"
/db_xref="MGI:1900711"
/db_xref="taxon:10090"
/clone="1110032P04"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18-day embryo"
42..1310
/notes="unannotated protein product; carboxypeptidase B2
(plasma) (MGI:1891837, GB|NM_019775, evidence: BLASTN,
9%, match=1419)
putative"
/codon_start=1
/protein_id="BAB23141.1"
/db_xref="GI:12835068"
CDS
source
FEATURES

```



```
/db_xref="WGI:1891837"
/translation="MKLHGLILVAIIILYEQHGFQSGQVLSALPRTSRQVOLLQNL
TTTIEVLMQFVTAETLEKKKEVFFVNASDVSVKALHNSRIPNVLNKNVEDLIE
QVFNDFVSPRASASYEQYHSLNEIYEWIEVTEQHPDMLQRIYTGSSPEPLPLVL
KVSKEQRIKNATWIDCGIHRANEWISPAFLWFIYVTOFHKENIYLRLLRHVDFYI
MPVMNVDDYDVTWKNRMWRKNRSAHKNRNCVGTDLNRNFPASKHCEKSGASSSCEST
YGLYDPESEBEVAVADFLRNIDHAKVISMHSYQOILPFPYSYNRSKSDHEELSL
VASENVAIRIESINKNRYTHSGSESILYAPGSDDDIYDLGKIYFTIELRTGRI
FLIPERYIKPTCAEALAAISKIVHVRNT"
BASE COUNT 421 a 329 c 311 g 385 t
ORIGIN

Query Match 41.78; Score 656; DB 11; Length 1446;
Best Local Similarity 72.2; Pred. No. 5.1e-70;
Matches 1009; Conservative 0; Mismatches 215; Indels 174; Gaps 5;

Qy 4 AAATGCTGTGGGATGAAGCTTTGCAGCCTTGCGAGCTTGAGTACCCATGTTCTCTCT 63
Db 28 AAGTCACTGTGGGATGAAGCTTCATGCGCTTGGAATCCTGTGATGCCATCATCTCTA-- 85
Qy 64 GTGAGCAGCATGCTTCGCGTTCAGAGTGGCCAGTGTCTAGTGTCTTCTTAGAACCT 123
Db 86 -TGAGCAGCATGCTTCGCGCTTCAGAGTGGCCAGTGTCTAGTGTCTTCTTAGAACCT 144
Qy 124 CTAGGCAAGTCAAGTCTACAGATCTTACTACACATATGAGATGTTCTCTGCGAGC 183
Db 145 CCAGGCAAGTCAAGTCTTACTACACATGATGAGTGTCTCTCTGCGAGC 204
Qy 184 CGGTAAACAGCTGACCTTATGTGAAGAAAAAACAAGTCCATTTTTTTGTAAGTGCATCTG 243
Db 205 CAGTCAAGCTGATTCATCGAGNAGNAGNAGTCCATTTTTTTGTAAGTGCATCTG 264
Qy 244 ATGTGCAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTGTGCG 303
Db 265 ATGTGCAAGTGTCAAGCGCATTTAAATGTGAGCAATTCATTTAAAGTCTGTATGA 324
Qy 304 CAGAGCTGGAAGATCTTATTCACAGCAGATTTCCACAGCAGCAGTCAAGCCCGGAGCCT 363
Db 325 ACAAGCTGGAAGATCTTATTCACAGCAGCAGATTTCCACAGCAGCAGTCAAGCCCGGAGCCT 384
Qy 364 CCGCATCTACTATGAACAGTATCACTCACTAAATGAATCTATCTTGGATAGAAATTA 423
Db 385 CCGCTTCACTATGACGATATCACTCGTAAATGAATCTATCTTGGATAGAAATTA 444
Qy 424 TAATGAGAGGATCTCTGATATGCTTACAAAAATCCACATTTGGATCTCTATTTGAGAAGT 483
Db 445 TAATGAGAGGATCTCTGATATGCTTCCAGAAATCTACATCGGATCATCATTCGAGAAGT 504
Qy 484 ACCCACTATATTTTAAAGTTCTGGAAGAAAGAAACAAAGCAAAATGCCATATGGA 543
Db 505 ACCCACTATATTTTAAAGTTCTGGAAGAAAGAAACAAAGCAAAATGCCATATGGA 564
Qy 544 TTGACTGTGGAATCCATGCGAGAGATGCTCTCTGCTTCTGCTTCTGCTGCTGCTATAG 603
Db 565 TCGACTGTGGAATCCATGCGAGAGATGCTCTCTGCTTCTGCTTCTGCTGCTGCTATAG 624
Qy 604 GC----- 605
Db 625 GCTACGTGACACATTCCTATGGGAAAGAAATCTGTATACAGACTTCTGAGGCACTGG 684
Qy 606 -----CATATC 612
Db 685 ATTTCTACATCATGCCCGTGTATGAACGCGATGCTATGACTACACGCGGAAAAAGAAATC 744
Qy 613 GAATGTGGAAGAAAGAAAGCTTTCTTATGCGACATCATTCATGCGGAGCAAGCACTGA 672
Db 745 GAATGTGGAAGAAAGAAAGCTTCTGCTCACAAGAAACAAACCGCTCGGTGGGACAGACCTGA 804
Qy 673 ATAGCAATTTTGTCTCAAAACACTGTGTGAGGAGGTGCATCCAGTTCCTCATGCTCGG 732
Db 805 ACAGGAATCTGCTTCCAAACACTGTGTGAGGAGGTGCATCCAGTTCCTCATGCTCGG 864
Qy 733 AAACCTACTGTGACTTTATCTCTGAGTCAGAACCAAGAGTGAAGGCGAGTGGCTAGTTTCT 792
```

```
Db 865 AAACCTACTGTGACTTTATCTCTGAGTCTGAGCAGAGGTGAAGGCGAGTGGCTACTTCT 924
Qy 793 TGAGAGAAATATCAACCCAGATTAAGAGCATACATCATGAGCATGCAATTCATATCTCCAGCATA 852
Db 925 TGAGAGAAATATCGACCAATTAAGCTTACATCATGATGCACTCATATCTCCCAACAA 984
Qy 853 TAGTGTTCATATTCCTATACACGAGTAAAGCAAGCAACCATGAGGAACCTGTCTCTAG 912
Db 985 TACTGTTCCTATTCCTATTAACAGAGCAAAAGCAAGCAACCATGAGGAACCTGTCTCTAG 1044
Qy 913 TACCCAGTGAAGCAGTTCGTGCTATTGACAAACTAGTAAATAACCCAGGTATACACATG 972
Db 1045 TGCCAGCAGCAAGCAGTTCGTGCAATTAAGAGTAAATAACCCAGGTATACACATG 1104
Qy 973 GCATGCTCAGAAACCTTATACCTAGTCTCCTGAGGTGGGACGATTCGATCTCATGAT 1032
Db 1105 GCATGCTCAGAAAGTTTATATCTAGTCTCTGAGGTCTGAGGATTCGATCTCATGAT 1164
Qy 1033 TGGGCATCAAAATATTCGTT----- 1051
Db 1165 TGGGCATCAAAATATTCGTTTACAATTCAGCTCCGAGATACAGCAGATACGGATTCCTGC 1224
Qy 1052 -----TACATCAACCCACCTGTAGAGAGCTTTTGGCGTGTCTTAAAAATAG 1100
Db 1225 TGCCAGAGATACATCAACCCACCTGTGAGAGAGCTTTTGGCGGCGCATCTTAAAAATAG 1284
Qy 1101 CTGCGCATGTCATTAGGAATGTTTAAATGCCCTGATTTTATCTTCTGCTTCCGATATTT 1160
Db 1285 TTTGGCATGTCATCAGAACATTAATGCC-----TAACCTCCGCTCTCATATTTT 1338
Qy 1161 AATTTACTGATCCAGCAAGCAACCAATCAATGTATCAGAT--TATTTTAAAGTTTATCC 1218
Db 1339 ATTTTATGATTTTCAACACATTAATGTTGATGCTTCTAAAGTTGAATCAGTTTCC 1398
Qy 1219 GTAGTTTGTATAAAGAT 1236
Db 1399 TTGTTTGTGGAAGAT 1416

RESULT 5
BG616950 780 bp mRNA linear EST 18-APR-2001
LOCUS 602615625F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4734247 5',
DEFINITION mRNA sequence.
ACCESSION BG616950
VERSION BG616950.1 GI:13668321
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 780)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Prepared by: The I.M.A.G.E. Consortium, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI597 row: b column: 08
High quality sequence stop: 754.
FEATURES
Location/Qualifiers
1..780
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4734247"
```



```
/lab host="DH10B (T1 phage-resistant)"
/clone lib="NIH_MGC_76"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 237 a 147 c 146 g 250 t
ORIGIN

Query Match 39.6%; Score 622.8; DB 10; Length 780;
Best Local Similarity 91.5%; Pred. No. 7.2e-66;
Matches 703; Conservative 0; Mismatches 12; Indels 53; Gaps 2;

Qy 850 ATATAGTGTTCCTATATTCCTATACAGAAAGTAAAGACCATGAGGAACTGTCTC 909
Db 1 ATATAGTGTTCCTATATTCCTATACAGAAAGTAAAGACCATGAGGAACTGTCTC 60
Qy 910 TAGTAGCAGTGAAGCAGTGTCTATTCACAAACTAGTAAATAATACCAGGTATACAC 969
Db 61 TAGTAGCAGTGAAGCAGTGTCTATTCAGAAATTAGTAAATAATACCAGGTATACAC 120
Qy 970 ATGCCATGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGAGCATTGGATCTATG 1029
Db 121 ATGCCATGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGAGCATTGGATCTATG 180
Qy 1030 ATTGGGCATCAATATTTCTG----- 1050
Db 181 ATTGGGCATCAATATTTCTGTTTCAATGAACTTCAGATACGGGCACATACGATCT 240
Qy 1051 -----TTACATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGTCTCTAAAA 1097
Db 241 TGCTGCGGAGCGTTACATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGTCTCTAAA 300
Qy 1098 TAGCTTGGCATGTCATTAGGATGTTTAAGTCCCTGATTTATCATCTGCTCCGTAT 1157
Db 301 TAGCTTGGCATGTCATTAGGATGTTTAAGTCCCTGATTTATCATCTGCTCCGTAT 360
Qy 1158 TTTAAATTTACTGATTCAGAGAACCAAAATCATTTGATCATCATTTTAAAGTTTATC 1217
Db 361 TTTAAATTTACTGATTCAGAGAACCAAAATCATTTGATCATCATTTTAAAGTTTATC 420
Qy 1218 CGTAGTTTTGATAAAAGATTTTCTATTTCTTGTGTTCTGTACAGAACCTAATAAGTGCT 1277
Db 421 CGTAGTTTTGATAAAAGATTTTCTATTTCTTGTGTTCTGTACAGAACCTAATAAGTGCT 480
Qy 1278 ACTTTGGCATTAAGCGAGATAGGTTTCATGCTTTTATACCTTTTAAAAAATAATTGTAA 1337
Db 481 ACTTTGGCATTAAGCGAGATAGGTTTCATGCTTTTATACCTTTT-AAAAAATAATTGTAA 539
Qy 1338 AAGCTAGTTACCTACTTTTCTTGTATTTTCGAGCTTTTGACTAGCCATCTCAGCACT 1397
Db 540 AAGCTAGTTACCTACTTTTCTTGTATTTTCGAGCTTTTGACTAGCCATCTCAGCACT 599
Qy 1398 TTCGACGTTTGATCAGCATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGTATCAT 1457
Db 600 TTCGACGTTTGATCAGCATCTCAAGCAAGTTTAAATCAAGATCATCTCAGCTGTATCAT 659
Qy 1458 TGGATCTCTACCAAAAGAAAGGGTGGTACAGATGATCAATTAAGATTTCTGCTCAAA 1517
Db 660 TGGATCTCTACCAAAAGAAAGGGTGGTACAGATGATCAATTAAGATTTCTGCTCAAA 719
Qy 1518 TTTTCAATTAATTTCTTCTCTCTTTTAAAAAATAAAAAAATAAAAAA 1565
Db 720 TTTTCAATTAATTTCTGTTGCTCTTAGTAAAAAATAAAAAAATAAAAAA 767

FEATURES
source
1..806
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4723690"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 239 a 158 c 162 g 247 t
ORIGIN

Query Match 39.1%; Score 615.6; DB 10; Length 806;
Best Local Similarity 90.7%; Pred. No. 5.2e-65;
Matches 724; Conservative 0; Mismatches 14; Indels 60; Gaps 4;

Qy 614 AATGTGGAGAAAGAACCGTTTCTTCTATGCGAAACAATCATTCATCGGAACAGACCTGAA 673
Db 1 AATGTGGAGAAAGAACCGTTTCTTCTATGCGAAACAATCATTCATCGGAACAGACCTGAA 60
Qy 674 TAGCAACTTGTCTCCAAACACTGGTGTAGGAGGTGATCCAGTTCTCTCATGCTCGGA 733
Db 61 TAGCAACTTGTCTTCCAAACACTGGTGTAGGAGGTGATCCAGTTCTCTCATGCTCGGA 120
Qy 734 AACCTACTGTGGACTTTTATCTCAGTCAGAACAGAGTGAAGCGCAGTGGCTAGTTCTT 793
Db 121 AACCTACTGTGGACTTTTATCTCAGTCAGAACAGAGTGAAGCGCAGTGGCTAGTTCTT 180
Qy 794 GAGAAGAAATATCAACAGATTTAAAGATACATCATCAGATGCATTCATCTCCAGCATAT 853
Db 181 GAGAAGAAATATCAACAGATTTAAAGATACATCATCAGATGCATTCATCTCCAGCATAT 240
Qy 854 AGTGTTCATATTCCTATACAGAGTAAAGCAAGCAATGAGAACTGCTCTAGT 913
Db 241 AGTGTTCATATTCCTATACAGAGTAAAGCAAGCAATGAGAACTGCTCTAGT 300
Qy 914 AGCCAGTAGAGCAGTTCGTCTATTGACAAAACCTAGTAAAAATACCAGGTATACATGG 973
```

```

Db 301 AGCCAGTGAACAGCTCGTGTATTGAGAAACTAGTAAATAACAGGTATACACATGG 360
Qy 974 CCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGGACGATTTGATGATTT 1033
Db 361 CCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGGACGATTTGATGATTT 420
Qy 1034 GGGCATCAATATATTCGTTT----- 1052
Db 421 GGGCATCAATATATTCGTTTCAATTTGAACCTTCGAGATACGGGACATACCGGATTTTGCT 480
Qy 1053 -----ACATCAAAACCCACCTGTAGAGAGCTTTTGGCCGCTCTCTCTAAATAGC 1101
Db 481 GCCGAGCGGTACATCAACCCACCTGTAGAGAGCTTTTGGCCGCTCTCTCTAAATAGC 540
Qy 1102 TTGGCATGTCTATAGGAATGTTTAATGCCCTGATTTTATCATCTCTGCTTCGGTATTTTA 1161
Db 541 TTGGCATGTCTATAGGAATGTTTAATGCCCTGATTTTATCATCTCTGCTTCGGTATTTTA 600
Qy 1162 ATTACTGATTTCCAGCAAGACCAATCATTTGATCATGATTTTAAAGTTTATCCGTA 1221
Db 601 ATTACTGATTTCCAGCAAGACCAATCATTTGATCATGATTTTAAAGTTTATCCGTA 660
Qy 1222 -GTTTTCATAAAGATTTTCTATTCTTCTGTTCTGTGAGAACCTTAATAAGTGTACT 1280
Db 661 GGTTCATAAAGATTTTCTATTCTTCTGTTCTGTGAGAACCTTAATAAGTGTAA-- 718
Qy 1281 TTGCATTAAGGAGACTAGGTTTCATGTTTATACCTTTTAAATAAAATTTGTAAG 1340
Db 719 -TTGCATTAAGGAGACTAGGTTTCATGTTTATACCTTTTAAATAAAATGT---AAAG 773
Qy 1341 TCTAGTTACCTACTTTT 1358
Db 774 TCTAGTTACTAAATTTCTT 791

```

```

RESULT 7
BG618629
LOCUS BG618629 788 bp mRNA linear EST 18-APR-2001
DEFINITION 602646064f1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
mRNA sequence.
ACCESSION BG618629
VERSION BG618629.1 GI:13670000
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1630 row: o column: 15
High quality sequence stop: 760.
Location/Qualifiers
1. 788
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4767974"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1:
SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor

```

FEATURES

source

```

RESULT 8
HSM089979
ID HSM089979 standard; RNA; EST; 702 BP.
XX
XX AC BX496754;
XX SV BX496754.1
XX
XX 09-MAY-2003 (Rel. 75, Created)
DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
XX Homo sapiens mRNA; EST DKFZp7911529_r1 (from clone DKFZp7911529)
XX
XX EST; expressed sequence tag.

```

sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGCGCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.95
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."

BASE COUNT 222 a 157 c 169 g 240 t

ORIGIN

```

Query Match 38.4%; Score 604.2; DB 10; Length 788;
Best Local Similarity 99.5%; Pred. No. 1.2e-63;
Matches 606; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AGAAATATGCTGTGGGATGAAGCTTTGACGCTTGACGCTTGTACCCATTGTTCTCT 60
Db 49 AGAAATATGCTGTGGGATGAAGCTTTGACGCTTGACGCTTGTACCCATTGTTCTCT 108
Qy 61 TCTGTGAGCAGCATGCTTTCGCGTTCCAGAGTGGCCAGTTCTAGCTCTTCTTAGAA 120
Db 109 TCTGTGAGCAGCATGCTTTCGCGTTCCAGAGTGGCCAGTTCTAGCTCTTCTTAGAA 168
Qy 121 CCTCTAGGCAAGTTCAAGTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGC 180
Db 169 CCTCTAGGCAAGTTCAAGTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGC 228
Qy 181 AGCCGGTAAACAGCTGACCTTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCA 240
Db 229 AGCCGGTAAACAGCTGACCTTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCA 288
Qy 241 CTGATGTGCAATGTGAAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 300
Db 289 CTGATGTGCAATGTGAAAGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGC 348
Qy 301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTTCCACACACAGTCAGCCCCGAG 360
Db 349 TGGCAGATGGAAGATCTTATTCAACAGCAGATTTTCCACACACAGTCAGCCCCGAG 408
Qy 361 CTCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGA 420
Db 409 CTCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGA 468
Qy 421 TTATAACTGAGAGGCATCTGTATATGCTTACAAAAATCCACATTCATTCATTTGAGA 480
Db 469 TTATAACTGAGAGGCATCTGTATATGCTTACAAAAATCCACATTCATTCATTTGAGA 528
Qy 481 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAAGCAAAACAGCCAAAATGCCAT 540
Db 529 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAAGCAAAACAGCCAAAATGCCAT 588
Qy 541 GGATTGACTGTGGAATCCATGCCAGAGATGGAATCTCTCTGCTTTCTGCTTGTGGTTCA 600
Db 589 GGATTGACTGTGGAATCCATGCCAGAGATGGAATCTCTCTGCTTTCTGCTTGTGGTTCA 648
Qy 601 TAGGCCATA 609
Db 649 TAGGCCATA 657

```

```
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP 1-702
RA Ansoerge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RT Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
XX This is the 5' sequence of the clone insert
XX Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
XX Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
XX sequenced by EMBL (European Molecular Biology Laboratories,
XX Heidelberg/Germany) within the cDNA sequencing consortium of
XX the German Genome Project.
XX No sl sequence available.
XX This clone (DKFZp77911529) is available at the RZPD in Berlin.
XX Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
XX 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
XX Key Location/Qualifiers
XX source
XX 1. 702
XX /db_xref="taxon:9606"
XX /mol_type="mRNA"
XX /organism="Homo sapiens"
XX /clone_lib="DKFZp77911529"
XX /clone_lib="779 (synonym: hnccl). Vector pSport1_sfi; host
XX DH10B; sites SfiI + SfiIb"
XX /dev_stage="fetal"
XX /tissue_type="liver"
XX
XX SQ Sequence 702 BP; 197 A; 149 C; 146 G; 210 T; 0 other;
XX
XX Query Match 37.8%; Score 594.8; DB 2; Length 702;
XX Best Local Similarity 99.5%; Pred. No. 1.8e-62;
XX Matches 607; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
XX
XX 1 AGAAATTTGCTTGGGATGAAGCTTTGCAGCCTTGAGCTCTGTACCCATTGTTCTCT 60
XX 42 AGAAATTTGCTTGGGATGAAGCTTTGCAGCCTTGAGCTCTGTACCCATTGTTCTCT 101
XX
XX 61 TCTGTGACGACATGTCTTCGCGTTCCAGAGTGCCCAAGTTCTAGCTCTTCCTAGAA 120
XX 102 TCTGTGACGACATGTCTTCGCGTTCCAGAGTGCCCAAGTTCTAGCTCTTCCTAGAA 161
XX
XX 121 CCTTAGGCAAGTTCAGTTCACAGATCTTACTACACATATGAGATTGTTCTCTGGC 180
XX 162 CCTTAGGCAAGTTCAGTTCACAGATCTTACTACACATATGAGATTGTTCTCTGGC 221
XX
XX 181 AGCCGGTAAACAGCTGACCTTATCTGTAAG-AAAAACAAGTCCATTTTGTGTAATGCA 239
XX 222 AGCCGGTAAACAGCTGACCTTATCTGTAAG-AAAAACAAGTCCATTTTGTGTAATGCA 281
XX
XX 240 TCTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCGTGTCTTG 299
XX 282 TCTGATGTCGACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCGTGTCTTG 341
XX
XX 300 CTGCGACAGTGGAGATCTTATTCACAGCAGATTTCCACGACACAGTCAGCCCCCGA 359
XX 342 CTGCGACAGTGGAGATCTTATTCACAGCAGATTTCCACGACACAGTCAGCCCCCGA 401
XX
XX 360 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTCTTGATAGAA 419
XX 402 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTCTTGATAGAA 461
XX
XX 420 TTTATACTGAGAGGCATCTGTATATGCTTACAAAAATCCACATTTGATCTCATTTTGG 479
XX 462 TTTATACTGAGAGGCATCTGTATATGCTTACAAAAATCCACATTTGATCTCATTTTGG 521
```

```
QY 480 AAGTACCACCTCTATGTTTAAAGGTTTCTGGAAAGACAAACAGCCAAAATGCCATA 539
DB 522 AAGTACCACCTCTATGTTTAAAGGTTTCTGGAAAGACAAACAGCCAAAATGCCATA 581
QY 540 TGGATTGACTGTGGAATCCATGCCAGAGATGGAATCTCTCTCTCTCTCTCTCTCTCT 599
DB 582 TGGATTGACTGTGGAATCCATGCCAGAGATGGAATCTCTCTCTCTCTCTCTCTCTCT 641
QY 600 ATAGGCCATA 609
DB 642 ATAGGCCATA 651

RESULT 9
BG566051
LOCUS
DEFINITION BG566051 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5',
mRNA sequence.
ACCESSION BG566051
VERSION BG566051.1 GI:13573704
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1549 row: 9 column: 13
High quality sequence stop: 731.
FEATURES
Location/Qualifiers
source
1..747
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/lab host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1:
SfiI (ggcgcctcgcc); Site 2: SfiI (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT 212 a 157 c 159 g 219 t
ORIGIN
Query Match 36.9%; Score 581.2; DB 10; Length 747;
Best Local Similarity 99.2%; Pred. No. 7.2e-61;
Matches 605; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
QY 1 AGAAATTTGCTTGGGATGAAGCTTTGCAGCCTTGAGCTCTGTACCCATTGTTCTCT 60
DB 4 AGAAATTTGCTTGGGATGAAGCTTTGCAGCCTTGAGCTCTGTACCCATTG-TCCT 62
QY 61 TCTGTGACGACATGTCTTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGTCTTCCTAGAA 120
DB 63 TCTGTGACGACATGTCTTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGTCTTCCTAGAA 122
```

Qy 121 CCTTAGGCAAGTTCAAGTTCTACAGAACTTACTACAAATATGAGATTGTTCTCTGGC 180
 Db 123 CCTTAGGCAAGTTCAAGTTCTACAGAACTTACTACAAATATGAGATTGTTCTCTGGC 182
 Qy 181 AGCGGTAACAGCTGACCTTATTGTGTAAGAAAAAACAAGTCCATTTTTTTGTAAATGTCAT 240
 Db 183 AGCGGTAACAGCTGACCTTATTGTGTAAGAAAAAACAAGTCCATTTTTTTGTAAATGTCAT 242
 Qy 241 CTGATGTCGAACAATGTGAAGGCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTTC 300
 Db 243 CTGATGTCGAACAATGTGAAGGCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTTC 302
 Qy 301 TGGCAGAGCTGGAGATCTTATTCAACAGCAGATTTTCAACGACACAGTCAAGCCCGGAG 360
 Db 303 TGGCAGAGCTGGAGATCTTATTCAACAGCAGATTTTCAACGACACAGTCAAGCCCGGAG 362
 Qy 361 CCTCCGATCGTACTATGAACAGTATCAGTCACTCAATGAAATCTATTCTTGGATAGAAT 420
 Db 363 CCTCCGATCGTACTATGAACAGTATCAGTCACTCAATGAAATCTATTCTTGGATAGAAT 422
 Qy 421 TTATAACTGAGAGGATCTCTGATATGCTTACAAAAATCCATTTGGATCTCAATTTGAGA 480
 Db 423 TTATAACTGAGAGGATCTCTGATATGCTTACAAAAATCCATTTGGATCTCAATTTGAGA 482
 Qy 481 AGTACCCACTCTATGTTT-TAAAGGTTTCTGGAAGAAACAACAGCCAAAAATGGCATA 539
 Db 483 AGTACCCACTCTATGTTTCTAAAGGTTTCTGGAAGAAACAACAGCCAAAAATGGCATA 542
 Qy 540 TGGATTGACGTGGAATCCATGCCAGAGATGATCTCTCTGTTCTGCTTTGCTGTTTC 599
 Db 543 TGGATTGACGTGGAATCCATGCCAGAGATGATCTCTCTGTTCTGCTTTGCTGTTTC 602
 Qy 600 ATAGGCCATA 609
 Db 603 ATAGGCCATA 612

RESULT 10

LOCUS BG618239
 DEFINITION 602645967P1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5',
 mRNA sequence.

ACCESSION BG618239

VERSION BG618239.1 GI:13669610

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE 1
 AUTHORS NIH-MGC
 TITLE NIH-MGC
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished

Contact: Robert Strauberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLCMI629 row: f column: 17

High quality sequence stop: 723.

Location/Qualifiers

1. .725

FEATURES

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4767376"

/lab_host="DHI08 (T1 phage-resistant)"

/clone_lib="NIH_MGC_76"

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctggcc); Site 2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCAGATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."

BASE COUNT 203 a 153 c 149 g 220 t
 ORIGIN

Query Match 36.8%; Score 579.2; DB 10; Length 725;

Best Local Similarity 99.2%; Pred. No. 1.3e-60;

Matches 603; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 1 AGAAAAATCTGTTGGGATGAAAGCTTTGAGCCCTTGCAGCTTCTGAGTCCCTGTGACCAATGTTCTCT 60

Db 48 AGAAAAATCTGTTGGGATGAAAGCTTTGAGCCCTTGCAGCTTCTGAGTCCCTGTGACCAATGTTCTCT 107

Qy 61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGATGCGCAAGTTCCTAGCTGCTCTTCTCTAGAA 120

Db 108 TCTGTGAGCAGCATGTCTTCGCGTTCCAGATGCGCAAGTTCCTAGCTGCTCTTCTCTAGAA 167

Qy 121 CCTTAGGCAAGTTCAAGTTCTACAGAACTTCTACTACAAATATGAGATTGTTCTCTGGC 180

Db 168 CCTTAGGCAAGTTCAAGTTCTACAGAACTTCTACTACAAATATGAGATTGTTCTCTGGC 226

Qy 181 AGCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGTCAT 240

Db 227 AGCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGTCAT 285

Qy 241 CTGATGTCGACAAATGTGAAGCCCAATTTAAATGTGAGCGGAATTCATGCAAGTCTTTC 300

Db 286 CTGATGTCGACAAATGTGAAGCCCAATTTAAATGTGAGCGGAATTCATGCAAGTCTTTC 345

Qy 301 TGGCAGAGCTGGAAGATCTTATTCAACAGCAGATTTTCAACGACACAGTCAAGCCCGGAG 360

Db 346 TGGCAGAGCTGGAAGATCTTATTCAACAGCAGATTTTCAACGACACAGTCAAGCCCGGAG 405

Qy 361 CCTCCGATCGTACTATGAACAGTATCAGTCACTCAATGAAATCTATTCTTGGATAGAAT 420

Db 406 CCTCCGATCGTACTATGAACAGTATCAGTCACTCAATGAAATCTATTCTTGGATAGAAT 465

Qy 421 TTATAACTGAGAGGCAATCTGATATGCTTACAAAAATCCACATTTGGATCTCTCAATTTGAGA 480

Db 466 TTATAACTGAGAGGCAATCTGATATGCTTACAAAAATCCACATTTGGATCTCTCAATTTGAGA 525

Qy 481 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAACAGCCAAAAATGGCATA 540

Db 526 AGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAACAGCCAAAAATGGCATA 585

Qy 541 GGATTGACGTGGAATCCATGCGCAGAGATGAGATCTCTCTGCTTCTGCTTGGTTCA 600

Db 586 GGATTGACGTGGAATCCATGCGCAGAGATGAGATCTCTCTGCTTCTGCTTGGTTCA 645

Qy 601 TAGGCCAT 608

Db 646 TAGGCCAT 653

RESULT 11

LOCUS

BG616456

DEFINITION

602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
 mRNA sequence.

ACCESSION BG616456

VERSION BG616456.1 GI:13667827

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

738 bp mRNA linear EST 18-APR-2001
 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
 mRNA sequence.

ACCESSION BG616456

VERSION BG616456.1 GI:13667827

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 738)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCMI595 row: c column: 05
 High quality sequence stop: 666.
 Location/Qualifiers
 1. 738
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4733500"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
 Sfil (ggccgctcgcc); Site 2: Sfil (ggccattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-ATTCTAGCGAGCGGCGGACATG-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."
 BASE COUNT 209 a 161 c 153 g 215 t
 ORIGIN
 Query Match 35.7%; Score 561.8; DB 10; Length 738;
 Best Local Similarity 99.6%; Pred. No. 1.5e-58;
 Matches 563; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 45 GTACCCATTGTTCTTTCTGTGAGCAGCATGCTTCGCGTTTCAGAGTGGCCAAAGTTCTA 104
 DB 1 GTACCCATTGTTCTTTCTGTGAGCAGCATGCTTCGCGTTTCAGAGTGGCCAAAGTTCTA 60
 QY 105 GCTGCTCTTCTAGAACTCTAGGCAAGTTCAAGTTCTCAGAACTCTTACTACAACTAT 164
 DB 61 GCTGCTCTTCTAGAACTCTAGGCAAGTTCAAGTTCTCAGAACTCTTACTACAACTAT 120
 QY 165 GAGATTGTTCTCTGCGAGCGGTAAACAGCTGACCTTATTTGTGAGAAACAAAGTCCAT 224
 DB 121 GAGATTGTTCTCTGCGAGCGGTAAACAGCTGACCTTATTTGTGAGAAACAAAGTCCAT 180
 QY 225 TTTTGTGTAATGATCTGATGTCGACAAATGTGAAGCCCACTTTAAATGTGAGCGGAAT 284
 DB 181 TTTTGTGTAATGATCTGATGTCGACAAATGTGAAGCCCACTTTAAATGTGAGCGGAAT 240
 QY 285 CCATGCAAGTCTTCTGCGAGAGCGTGAAGATCTTATTAACAGAGATTTCCAAAGCAG 344
 DB 241 CCATGCAAGTCTTCTGCGAGAGCGTGAAGATCTTATTAACAGAGATTTCCAAAGCAG 300
 QY 345 ACAGTCAGCCCGGAGCTCGGATCGTACTATCAACAGATATCACTCACTAAATGAATC 404
 DB 301 ACAGTCAGCCCGGAGCTCGGATCGTACTATCAACAGATATCACTCACTAAATGAATC 360
 QY 405 TATTCTTGGATAGAAATTTATAACTGAGAGGATCTCTGATATGCTTACAAAATCCACAT 464
 DB 361 TATTCTTGGATAGAAATTTATAACTGAGAGGATCTCTGATATGCTTACAAAATCCACAT 420
 QY 465 GGATCCTCATTTGAGAGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAAGCAAAACA 524
 DB 421 GGATCCTCATTTGAGAGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAAGCAAAACA 480

QY 525 GCCAAAATGCCATATGATGATTGACTGTGAATCCATCCAGAGATGGATCTCTCTGCT 584
 DB 481 GCCAAAATGCCATATGATGATTGACTGTGAATCCATCCAGAGATGGATCTCTCTGCT 540
 QY 585 TTCTGCTGTGTGTTTCATAGCCATA 609
 DB 541 TTCTGCTGTGTGTTTCATAGCCATA 565
 RESULT 12
 AV646979 735 bp mRNA linear EST 15-JAN-2002
 LOCUS AV646979 GLC Homo sapiens cDNA clone GLCATD06 3', mRNA sequence.
 DEFINITION AV646979
 ACCESSION AV646979
 VERSION AV646979.1 GI:9867993
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 735)
 AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z. and Han, Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 PUBMED 11752456
 COMMENT Contact: Zequan Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919 (ex.45)
 Fax: 86-21-50801922
 Email: hanzq@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers
 1. 735
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCATD06"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 200 a 158 c 172 g 205 t
 ORIGIN
 Query Match 35.3%; Score 555.2; DB 9; Length 735;
 Best Local Similarity 98.4%; Pred. No. 9.6e-58;
 Matches 571; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
 QY 31 GCCTTCAGTCTTGTACCCATTGTTCTCTTGTGAGCAGCATGCTTCGCGTTCCAGA 90
 DB 5 GCCTTCGCGTCTTGTACCCATTGTTCTCTTGTGAGCAGCATGCTTCGCGTTCCAGA 64
 QY 91 GTGGCCAAAGTCTAGTGTCTTCTTAGACCTCTAGGCAAGTTCAGTTCTACAGATC 150
 DB 65 GTGGCCAAAGTCTAGTGTCTTCTTAGACCTCTAGGCAAGTTCAGTTCTACAGATC 124
 QY 151 TTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGA 210
 DB 125 TTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGA 184
 QY 211 AAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGCGCAATGTGAAGCCCATTTAA 270
 DB 185 AAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGCGCAATGTGAAGCCCATTTAA 244

QY 271 ATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTGGAGATCTTATTCAACAGC 330
 |||||
 Db 245 ATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTGGAGATCTTATTCAACAGC 304
 |||||
 QY 331 AGATTTCACACGACACAGTCCAGCCCGAGCCTCCGCATCGTACTATGAACAGTATCACT 390
 |||||
 Db 305 AGATTTCACACGACACAGTCCAGCCCGAGCCTCCGCATCGTACTATGAACAGTATCACT 364
 |||||
 QY 391 CACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCATCTCTGATATGCTTA 450
 |||||
 Db 365 CACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCATCTCTGATATGCTTA 424
 |||||
 QY 451 CAAAAATCCATCTGATCTCTTGGAGAGTACCACCTCTATGTTTAAA-GGTTTCT 509
 |||||
 Db 425 CAAAAATCCATCTGATCTCTTGGAGAGTACCACCTCTATGTTTAAAAGGGTGTCT 484
 |||||
 QY 510 GGAAGAGAACAAACAGCAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
 |||||
 Db 485 GGAAGAGAACAAACAGCAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 544
 |||||
 QY 570 TGGATCTCTCTGCTGTTCTGTTGTTGTTTCATAGGCCATA 609
 |||||
 Db 545 TGGATCTCTCTGCTGTTCTGTTGTTGTTTCATAGGCCATA 584
 |||||

RESULT 13

CB157619
 LOCUS K-EST0216807 L17N670205n1 Homo sapiens cDNA clone EST 29-JAN-2003
 DEFINITION L17N670205n1-12-A11 5', mRNA sequence.
 ACCESSION CB157619
 VERSION CB157619.1 GI:28142751
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 541)
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongseung@mail.kribb.re.kr
 Plate: 12 row: A column: 11
 High quality sequence stop: 541.
 Location/Qualifiers

FEATURES

source
 1. .541
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="L17N670205n1-12-A11"
 /sex="F"
 /lab_host="Top10F"
 /clone_lib="L17N670205n1"
 /notes="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;
 Site 2: NotI; The library was contributed by the Soares
 laboratory and it was constructed as described by Ronaldo,
 M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
 6(9): 791-806. RNA was prepared from harvested cell
 culture."
 BASE COUNT 156 a 121 c 107 g 157 t

ORIGIN

Query Match 34.1%; Score 536.2; DB 14; Length 541;
 Best Local Similarity 99.4%; Pred. No. 2.2e-55;

Matches 538; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 22 AGCTTTGCAGCCTTGCAGTCTTGTATCCCATTTCTTCTTCTGTGAGCAGCATGTCTTCG 81
 |||||
 Db 1 AGCTTTGCAGCCTTGCAGTCTTGTATCCCATTTCTTCTTCTGTGAGCAGCATGTCTTCG 60
 |||||
 QY 82 CGTTCCAGAGTGGCCAAAGTCTTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAGATTC 141
 |||||
 Db 61 CGTTTCAGAGTGGCCAAAGTCTTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAGATTC 120
 |||||
 QY 142 TACAGAAATCTTACTACAAACATATGAGATTTGTTCTCTGGCAGCGGTAAACAGCTGACCTTA 201
 |||||
 Db 121 TACAGAAATCTTACTACAAACATATGAGATTTGTTCTCTGGCAGCGGTAAACAGCTGACCTTA 180
 |||||
 QY 202 TTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTGACAATGTGAAAG 261
 |||||
 Db 181 TTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTGACAATGTGAAAG 240
 |||||
 QY 262 CCATTTAAATGTGAGCGGAATTCATGCAATGTTCTTGTGGCAGAGCTGGGAAGATCTTA 321
 |||||
 Db 241 CCATTTAAATGTGAGCGGAATTCATGCAATGTTCTTGTGGCAGAGCTGGGAAGATCTTA 300
 |||||
 QY 322 TTCAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAGCCTCCGCATCGTACTATGAAC 381
 |||||
 Db 301 TTCAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAGCCTCCGCATCGTACTATGAAC 360
 |||||
 QY 382 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCATCTTG 441
 |||||
 Db 361 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCATCTTG 420
 |||||
 QY 442 ATATGCTTACAAAAATCCCATTTGGATGCTCTCATTTGAGAAGTACCCACTCTATGTTTAA 501
 |||||
 Db 421 ATATGCTTACAAAAATCCCATTTGGATGCTCTCATTTGAGAAGTACCCACTCTATGTTTAA 480
 |||||
 QY 502 AGGTTTCTGGAAGAACAAACAGCCCAAAATGCCATATGATGTTGACTGTGGAATCCATG 561
 |||||
 Db 481 AGGTTTCTGGAAGAACAAACAGCCCAAAATGCCATATGATGTTGACTGTGGAATCCATG 540
 |||||
 QY 562 C 562
 Db 541 C 541

RESULT 14

BG618813
 LOCUS 602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767885 5',
 DEFINITION mRNA sequence.
 ACCESSION BG618813
 VERSION BG618813.1 GI:13670184
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 750)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1630 row: k column: 22
 High quality sequence stop: 622.
 Location/Qualifiers

FEATURES

source
 1. .750
 /organism="Homo sapiens"

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4767885"
/lab_hosts="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattggcc); Site 2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCGCATATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCGGACATG-dt (30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT      219 a 146 c 171 g 213 t 1 others
ORIGIN
Query Match      33.8%; Score 532.2; DB 10; Length 750;
Best Local Similarity 98.0%; Pred. No. 5.4e-55;
Matches 581; Conservative 0; Mismatches 8; Indels 4; Gaps 4;
Qy 9 GCTGTTGGGATGAAGCTTTGCAGCTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGG 128
Db 2 GCTGTTGGGATGAAGCTTTGCAGCTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGG 61
Qy 69 CAGCATCTCTTCGGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGG 128
Db 62 CAGCATCTCTTCGGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGG 121
Qy 129 CAAGTTCAGTTCACAGATCTTACTACAACTATGAGATTTGTTCTCTGCGACCGGTA 188
Db 122 CAAGTTCAGTTCACAGATCTTACTACAACTATGAGATTTGTTCTCTGCGACCGGTA 181
Qy 189 ACAGCTGACCTTATTTGTAAGAAAAAACAAGTCCATTTTGTGAAATGCACTGTATGTC 248
Db 182 ACAGCTGACCTTATTTGTAAGAAAAAACAAGTCCATTTTGTGAAATGCACTGTATGTC 241
Qy 249 GACATGTGAAGCCCAATTTAAATGTGAGGGAATTCATGCACTGCTGCTGCGAGAC 308
Db 242 GACATGTGAAGCCCAATTTAAATGTGAGGGAATTCATGCACTGCTGCTGCGAGAC 301
Qy 309 GTGAAGATCTTATTAACAGCAGATTTCCAAAGACAGTCAGCCCCCGAGCTCCGCA 368
Db 302 GTGAAGATCTTATTAACAGCAGATTTCCAAAGACAGTCAGCCCCCGAGCTCCGCA 361
Qy 369 TCGTACTATGAACAGTATCACTCACTAAATGAAATCTTATTTCTGGATAGAAATTTATAACT 428
Db 362 TCGTACTATGAACAGTATCACTCACTAAATGAAATCTTATTTCTGGATAGAAATTTATAACT 421
Qy 429 GAGAGGATCTGATATGCTT-ACAAAATCCACATGG-ATCTCAATTTGAGAGTACC 486
Db 422 GAGAGGATCTGATATGCTTAAACAAAATCCACATTTGGCATCTCTCAATTCGAGAAGTAAAC 481
Qy 487 CACTC-TATGTTTAAAGTTTCT-GGAAAAGACAAACAGCCAAATGCGATATGAT 544
Db 482 CAGTCTGATGTTTAAAGTTTCTTGGGAAAGAAACAGAGGCAAAATGCCATATGAT 541
Qy 545 TGACTGTGGAATCCATGCCAGAGAAATGATCTCTCTGCTTTCTGCTGTGGT 597
Db 542 TGACTGTGGAATCCATGCCAGAGAAATGATCTCTCTGCTTTCTGCTGTGGT 594
```

```
RESULT 15
BG568240
LOCUS      602587149f1 NIH_MGC_76 Homo sapiens cdna clone IMAGE:4715797 5',
DEFINITION mRNA sequence.
ACCESSION BG568240
VERSION    BG568240.1 GI:13575893
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNCMI563 row: a column: 14
High quality sequence stop: 691.

FEATURES

source

1..691
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4715797"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattggcc); Site 2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCGCATATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGCGGCGGACATG-dt (30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT 204 a 143 c 144 g 200 t
ORIGIN

Query Match 33.3%; Score 523.8; DB 10; Length 691;
Best Local Similarity 98.5%; Pred. No. 5.8e-54;
Matches 539; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Qy 3 AAAATGCTGCTTGGGATGAAGCTTTGCAGCTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAAC 122
Db 4 AAAATGCTGCTTGGGATGAAGCTTTGCAGCTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAAC 123
Qy 63 TGTGAGCAGCATGCTCTTCGGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAAC 122
Db 64 TGTGAGCAGCATGCTCTTCGGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAAC 123
Qy 123 TCTAGGCAAGTTCAGATCTTACAGATCTTACTACAACTATGAGATTTGTTCTCTGGCAG 182
Db 124 TCTAGGCAAGTTCAGATCTTACAGATCTTACTACAACTATGAGATTTGTTCTCTGGCAG 183
Qy 183 CCGGTAAACAGTCACTTATTTGAGAAAAAACAAGTCCATTTTGTGAAATGCACT 242
Db 184 CCGGTAAACAGTCACTTATTTGAGAAAAAACAAGTCCATTTTGTGAAATGCACT 242
Qy 243 GATGTCGACAAATGGAAGCCCAATTTAAATGTGAGCGGAATTCATGCACTGCTGCTG 302
Db 243 GATGTCGACAAATGGAAGCCCAATTTAAATGTGAGCGGAATTCATGCACTGCTGCTG 302
Qy 303 GCAGAGTGAAGATCTTATTCAACAGCAGATTTCCAAACAGACAGTCAGCCCCCGAGCC 362
Db 303 GCAGAGTGAAGATCTTATTCAACAGCAGATTTCCAAACAGACAGTCAGCCCCCGAGCC 362
Qy 363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTCGATAGAAATTT 422
Db 363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTT 422
Qy 423 ATAACTGAGAGGATCTCTGATATGCTTTACAAAAATCCACATTTGGATCTCTATTGAGAAG 482
Db 423 ATAACTGAGAGGATCTCTGATATGCTTTACAAAAATCCACATTTGGATCTCTATTGAGAAG 482

Qy	483	TACCACTCTATGTTTAAAGGTTTCTGGAAAGAACAAACAGCCAAAATGCCATATGG	542
Db	483	TACCACTCTATGTTTAAAGGTTTCTGGAAAGAACAAACAGCCAAAATGCCATATAA	542
Qy	543	ATTGACT	549
Db	543	ACTCAAT	549

Search completed: January 1, 2004, 15:37:43
Job time : 3649.93 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:38:01 ; Search time 141.462 Seconds
(without alignments)
2423.606 Million cell updates/sec

Title: US-09-980-881a-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 agaaagcttgcagccttgc.....ttttgcccgtgtctctataaa 1080

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09980881/runat_31122003_135747_24935/app_query.fasta_1.2958
-DB=A_Geneseq_19Jun03_QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN 1.117 @runat_31122003_135747_24935 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- A_Geneseq_19Jun03:**
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:**
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:**
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:**
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:**
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:**
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:**
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:**
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:**
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:**
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:**
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:**
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:**
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:**
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:**
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:**
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:**
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:**
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:**
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:**
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:**
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:**
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:**
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:**
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	1911	98.9	360	21	AAB11457	Human brain carbox
2	1857.5	96.1	386	24	ABG72095	Human protease pro
3	1852.5	95.9	386	24	ABG72093	Human protease pro
4	1835	95.0	423	17	AAR90293	Human plasma carbo
5	1835	95.0	423	18	AAW14733	Human plasma carbo
6	1835	95.0	423	20	AAW22270	Human plasma carbo
7	1830	94.7	423	14	AAR36273	Human plasma carbo
8	1780	92.1	338	21	AAB11458	Human brain carbox
9	1320	68.3	246	21	AAB11459	Human brain carbox
10	839	43.4	211	21	AAB58129	Lung cancer associ
11	641	33.2	417	24	AB996688	Amino acid sequenc
12	640.5	33.2	404	16	AAR75131	Porcine Tyr-His-Me
13	639.5	33.1	415	17	AAW06172	Human pancreatic c
14	639.5	33.1	415	19	AAW74476	Human pancreatic c
15	636	32.9	324	22	AA666562	Bothops jararaca c
16	635	32.9	417	23	AAU84325	Protein CPA3 diffe
17	634	32.8	667	24	ABP97414	Anti-human seminal
18	632	32.7	424	17	AAW06175	ProHCPB with PelB
19	623	32.2	716	18	AAW41414	PreproHCPB-linker-
20	613.5	31.8	434	24	AB996687	Amino acid sequenc
21	613.5	31.8	437	22	AA800502	Human carboxypepti
22	613.5	31.8	437	23	AAE19175	Human protease, PR
23	604	31.3	307	17	AAW0602	Rat carboxypeptida
24	576	29.8	306	16	AAR75132	Porcine carboxypep
25	576	29.8	306	22	AAU04477	Porcine carboxypep
26	564.5	29.2	354	22	AA666560	Human secreted met
27	564.5	29.2	354	22	AA666566	Human secreted met
28	564.5	29.2	362	22	AA674682	Human protease and
29	564.5	29.2	374	22	AA666547	Human secreted met
30	564.5	29.2	374	22	AA666565	Human secreted met
31	564.5	29.2	374	23	AAU82703	Amino acid sequenc
32	564.5	29.2	374	24	ABJ37888	NOVX protein sequen
33	564.5	29.2	374	24	ABJ374658	Human zinc-binding
34	563	29.1	444	23	AAE24240	Human 23566 (carbo
35	561	29.0	329	17	AAW06174	Mature HCPB with P
36	561	29.0	349	17	AAW06173	mature HCPB (His)6
37	560	29.0	613	18	AAW41412	Humanised Pd-mutan
38	560	29.0	613	19	AAW74481	Human pancreatic c
39	554	28.7	349	18	AAW13749	Modified HCPB (D25
40	554	28.7	349	17	AAW13749	Carboxypeptidase B
41	553	28.6	349	17	AAW06182	Modified HCPB (D25
42	553	28.6	349	18	AAW13750	Carboxypeptidase B
43	550	28.5	349	18	AAW13763	Carboxypeptidase B
44	548.5	28.4	374	24	ABJ37887	NOVX protein sequen
45	548	28.4	349	18	AAW13752	Carboxypeptidase B

ALIGNMENTS

RESULT 1

AAB11457

ID AAB11457 standard; Protein; 360 AA.

AC AAB11457;

XX 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B protein.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; antialzheimers; nootropic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma.

XX Homo sapiens.

XX

cancer; arteriosclerosis; degenerative disorder; chromosome 13; single nucleotide polymorphism; SNP.

Homo sapiens.

Key	Location/Qualifiers
1	1.1
2	2.1
3	3.1
4	4.1
5	5.1
6	6.1
7	7.1
8	8.1
9	9.1
10	10.1
11	11.1
12	12.1
13	13.1
14	14.1
15	15.1
16	16.1
17	17.1
18	18.1
19	19.1
20	20.1
21	21.1
22	22.1
23	23.1
24	24.1
25	25.1
26	26.1
27	27.1
28	28.1
29	29.1
30	30.1
31	31.1
32	32.1
33	33.1
34	34.1
35	35.1
36	36.1
37	37.1
38	38.1
39	39.1
40	40.1
41	41.1
42	42.1
43	43.1
44	44.1
45	45.1
46	46.1
47	47.1
48	48.1
49	49.1
50	50.1
51	51.1
52	52.1
53	53.1
54	54.1
55	55.1
56	56.1
57	57.1
58	58.1
59	59.1
60	60.1
61	61.1
62	62.1
63	63.1
64	64.1
65	65.1
66	66.1
67	67.1
68	68.1
69	69.1
70	70.1
71	71.1
72	72.1
73	73.1
74	74.1
75	75.1
76	76.1
77	77.1
78	78.1
79	79.1
80	80.1
81	81.1
82	82.1
83	83.1
84	84.1
85	85.1
86	86.1
87	87.1
88	88.1
89	89.1
90	90.1
91	91.1
92	92.1
93	93.1
94	94.1
95	95.1
96	96.1
97	97.1
98	98.1
99	99.1
100	100.1

Misc-difference 169

/note= "This residue is the major variant of a single nucleotide polymorphism. The minor variant is an Thr residue"

US2002137179-A1.

26-SEP-2002.

21-MAR-2001; 2001US-0813133.

21-MAR-2001: 2001US-0813133.

(GANW/) GAN W.

(KETCHUM K A.

(DEFA/) DI FRANCESCO V.

(BEAS/) BEASLEY E M.

Gan W, Ketchum KA, Di Francesco V, Beasley EM;

WPI; 2003-102387/09.

N-PSDB; ABX13670, ABX13671.

New isolated human protease proteins, useful for developing therapeutic or diagnostic compositions, particularly for developing human therapeutic agents that modulate protease activity in cells or tissues

Claim 1: Fig 2: 80pp: English.

The invention discloses an isolated human protease peptide, its allelic variant or orthologue. The proper functioning of the cell requires careful control of the levels of important structural proteins, enzymes and regulatory proteins. One of the ways the cell reduces the steady state level of a particular protein is by proteolytic degradation. Proteolysis can also be used to convert a pre or pro-protein in to an active form. Proteases also regulate many different cell proliferation, differentiation and signalling processes. The peptides and nucleic acid molecules are useful in the development of human therapeutics (gene and protein therapy) and diagnostic compositions. The peptides are also useful for raising antibodies or eliciting an immune response (vaccine), as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids, or as markers for tissues in which the corresponding protein is preferentially expressed and in methods for identifying a modulator of the peptide or an agent that binds to the peptide. The agents identified are useful for treating protease-related conditions that are specific for the subfamily of proteases that the peptide belongs to, particularly in cells and tissues that express the protease, such as inflammation, cancer, arteriosclerosis and degenerative disorders. The modulator of the peptide is also useful for treating a disorder characterised by an absence of, inappropriate or unwanted expression of the protein. The sequence presented is the human protease protein, variant #1, the gene for which is located on chromosome 13.

Sequence 386 AA;

Alignment Scores:

Pred. No.:	1 57e-191	Length:	386
Score:	1852.50	Matches:	354
Percent Similarity:	94.16%	Conservative:	1
Best Local Similarity:	93.90%	Mismatches:	5
Query Match:	95.89%	Indels:	18
DB:	24	Gaps:	1

US-09-980-881A-1 COPY 18 1097 (1-1080) x ABG72093 (1-386)

Qy	1	ATGAGC	TTTGGAGC	CTTGGAGT	CTCTGT	TACCAT	TGTTCT	CTCTGT	GAGCAG	CAATGTC	60												
Db	1	MetLys	Leu	Cys	Ser	Leu	Ala	Val	Leu	Val	Pro	Ile	Val	Leu	Phe	Cys	Glu	Gln	His	Val	20		
Qy	61	TTCCG	TTTCAG	AGTGGC	CAAGTTC	TAGT	GTCT	CTCT	TAGAAC	CTCTAGG	CAAC	CTT	AGG	CAAGTTC	CAA	120							
Db	21	Phe	Ala	Phe	Gln	Ser	Gly	Gln	Val	Leu	Ala	Leu	Pro	Arg	Thr	Ser	Arg	Gln	Val	Gln	40		
Qy	121	GTCT	CTAC	AGAA	CTTACT	CAAC	ATAT	GAGAT	TGTTCT	CTGG	CAGCGG	GTAA	CAGCT	GAC	180								
Db	41	Val	Leu	Gln	Asn	Leu	Thr	Thr	Thr	Tyr	Glu	Ile	Val	Leu	Trp	Gln	Pro	Val	Thr	Ala	Asp	60	
Qy	181	CTTAT	TGTGA	AAAA	CAAGTCC	ATTTT	TGT	TAAAT	GCAT	CTCAT	GAT	TCG	CAAT	GTG	240								
Db	61	Leu	Ile	Val	Lys	Lys	Lys	Gln	Val	His	Phe	Phe	Val	Asn	Ala	Ser	Asp	Val	Asp	Asn	Val	80	
Qy	241	AAAGC	CCATTA	ATGT	GAGCG	CAAT	TCCAT	GCAGT	GTCT	TGCT	GCGAC	GC	TGG	AGAT	300								
Db	81	Lys	Ala	His	Leu	Asn	Val	Ser	Gly	Ile	Pro	Cys	Ser	Val	Leu	Leu	Ala	Asp	Val	Glu	Asp	100	
Qy	301	CTTAT	TCAAC	AGAG	ATTTT	CAAC	AGC	ACAC	AGT	CAG	CGCC	CCG	AGC	CTT	360								
Db	101	Leu	Ile	Gln	Gln	Gln	Ile	Ser	Asn	Asp	Thr	Val	Ser	Pro	Arg	Ala	Ser	Ala	Ser	Tyr	Tyr	120	
Qy	361	GAA	CAGT	ATC	ACT	CAATA	GAAT	CTAT	TTT	TGGAT	AGAA	TTT	TAA	CT	420								
Db	121	Glu	Gln	Tyr	His	Ser	Leu	Asn	Glu	Ile	Tyr	Ser	Trp	Ile	Glu	Phe	Ile	Thr	Glu	Ala	Ser	His	140
Qy	421	CCT	GATAT	GC	TTAC	AAAA	TCC	AT	TGGAT	TGAT	TGAG	AGT	TAC	CACT	480								
Db	141	Pro	Asp	Met	Leu	Thr	Lys	Ile	His	Ile	Gly	Ser	Ser	Phe	Glu	Lys	Tyr	Pro	Leu	Tyr	Val	160	
Qy	481	TTAA	AGT	TTCT	CGAAA	GAAG	CAAA	CAGC	CAAAA	ATG	CCCAT	TAT	GGA	TGAT	540								
Db	161	Leu	Lys	Val	Ser	Gly	Lys	Glu	Gln	Ala	Ala	Lys	Asn	Ala	Ile	Trp	Ile	Asp	Cys	Gly	Ile	180	
Qy	541	CAT	CC	CAG	AGA	TGG	AT	CTCT	CTG	CTTT	CTG	CTT	GTG	TT	600								
Db	181	His	Ala	Arg	Glu	Trp	Ile	Ser	Pro	Ala	Phe	Cys	Leu	Trp	Phe	Ile	Gly	His	Asn	Arg	Met	200	
Qy	601	TG	GAA	AA	GAAC	CGT	TTCT	TCT	ATG	CGA	CAAT	CGAT	TGC	GA	660								
Db	201	Trp	Arg	Lys	Asn	Arg	Ser	Phe	Tyr	Ala	Asn	Asn	His	Cys	Ile	Gly	Thr	Asp	Leu	Asn	Arg	220	
Qy	661	AAC	TTT	GTCT	CC	AAAA	CAC	TGCT	GTG	AGG	AGT	GGC	AT	CCAGT	720								
Db	221	Asn	Phe	Ala	Ser	Lys	His	Trp	Cys	Glu	Glu	Gly	Ala	Ser	Ser	Ser	Ser	Cys	Ser	Glu	Thr	240	
Qy	721	TAC	TG	TGG	ACT	TTT	CT	CTG	AGT	CAG	AA	CC	AGA	AGT	800								
Db	241	Tyr	Cys	Gly	Leu	Tyr	Pro	Glu	Ser	Glu	Pro	Glu	Val	Lys	Ala	Val	Ala	Ser	Phe	Leu	Arg	260	
Qy	781	AGA	AA	TAT	CAAC	AGAT	TAA	AGC	ATAC	AT	CAG	AT	GCAT	TCT	840								
Db	261	Arg	Asn	Ile	Asn	Gln	Ile	Lys	Ala	Tyr	Ile	Ser	Met	His	Ser	Tyr	Ser	Gln	His	Ile	Val	280	
Qy	841	TTT	CCAT	TAT	CT	CAT	CAC	GAAGT	TAA	AGC	CAAG	CCAT	CAG	AA	900								
Db	281	Phe	Pro	Tyr	Ser	Tyr	Thr	Arg	Ser	Lys	Ser	Lys	Asp	His	Glu	Glu	Leu	Ser	Leu	Val	Ala	300	
Qy	901	AGT	GAAG	CAG	CT	CGT	GTCT	TAT	TGA	CAAA	CACT	AGT	TAA	AAAA	960								
Db																							

||||| 361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaValSerLys 377
|||||

RESULT 4

AAR90293
ID AAR90293 standard; Protein; 423 AA.

AC AAR90293;

XX 25-MAR-2003 (updated)
DT 12-APR-1996 (first entry)

XX Human plasma carboxypeptidase B.

XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Active-site 181
/note= "Residue suspected of being involved in
catalytic activity of enzyme."

FT Binding-site 183
/note= "Residue suspected of being part of the
binding site of enzyme."

FT Binding-site 256..257
/note= "Residues suspected of being part of the
binding site of enzyme."

FT Active-site 310
/note= "Residue suspected of being involved in
catalytic activity of enzyme."

FT Binding-site 312
/note= "Residue suspected of being part of the
binding site of enzyme."

FT Active-site 369
/note= "Residue suspected of being involved in
catalytic activity of enzyme."

FT Binding-site 370..371
/note= "Residues suspected of being part of the
binding site of enzyme."

FT Binding-site 394
/note= "Residue suspected of being part of the
binding site of enzyme."

XX US5474901-A.

PD 12-DEC-1995.

XX 19-JUL-1994; 94US-0277540.

XX 01-FEB-1991; 91US-0649591.

PR 14-OCT-1992; 92US-0959944.

PR 13-DEC-1993; 93US-0167727.

PR 19-JUL-1994; 94US-0277540.

XX (GETH) GENENTECH INC.

XX Drayna DT, Eaton DL;

XX WPI; 1996-039508/04.

XX N-PSDB; AAT11671.

PT Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A

PS Disclosure; Figure 4; 40pp; English.

XX An antibody which specifically binds human plasma carboxypeptidase B
XX (hPCPB) and does not cross react with other carboxypeptidases is
CC useful for the detection of hPCPB in vitro. The antibody is also
CC used for purifying hPCPB from a sample. Purification comprises

CC passing a sample thought to contain hPCPB over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hPCPB.
CC (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 423 AA;

Alignment Scores:
Pred. No.: 1.29e-189 Length: 423
Score: 1835.00 Matches: 356
Percent Similarity: 86.23% Conservatives: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 17 Gaps: 2

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAR90293 (1-423)

QY 1 ATGAAGCTTTGAGCCCTTGAGTCCTTGACCCATTGTTCTCTCTCTGTGAGCAGCATGTC 60
DB 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 61 TTCGCGTTCAGAGTGCACCAAGTCTAGCTGCTCTTCTAGAACCTCTAGCAAGTTCAA 120
DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 121 GTTCTACAGAACTCTTACTACACATATGAGATTGTTCTCTGGCAGCCGCGTAACAGCTGAC 180
DB 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAAATGTG 240
DB 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 241 AAGGCCATTTAAATGTGAGCGGAATCCATGTCAGTCTCTTCTGGCAGACGCTGGAAGAT 300
DB 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY 301 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAGCCTCCGCATCGTACTAT 360
DB 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
QY 361 GAACAGTATCCTCCTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 420
DB 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 421 CTTGATATGCTTACAAAATCCACATGGATCTCTCATTTGAGAGTACCCACTATGTT 480
DB 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 481 TTAAGGTTTCTGAAAAGAAACAAACAGCCAAAATGCCATATGGATTGACTGTGGAATC 540
DB 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 541 CATGCCAGAGAATGATCTCTCTCTGCTTCTGCTTGTGTTTCATAGGCCAT----- 591
DB 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 591 ----- 591
DB 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 592 -----AATCGAATGTGAGAAAG 609
DB 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 610 AACCGTTCTTTCTATCGAACAAATATTGTCATCGAACACACCTGAATAGCAACTTTGTC 669
DB 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 670 TCAAAACACTGGTGTGAGGAGGTGATCCAGTTCCTCATCTGCTCGGAAACCTACTGTGGA 729
DB 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerSerSerSerSerSerSerSer 280

```

QY 730 CTTTATCTGAGTCAGAACAGAGTGAAGSCAGTGTCTTCTTGAGAGAAATATC 789
Db 281 LeuTyrProGluSerGluProGluValysAlaValAlaSerPheLeuArgGAsnIle 300
QY 790 AACGAGATTAAAGCATACATCAGCATGATTCATCTCCAGCATATAGTGTTCATAT 849
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 850 TCCTATACAGAGTAAGAAAGACAGCATGAGAACTGTCTAGTAGCCAGTGAAGCA 909
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
QY 910 GTTCGCTGCTATTGACAAACTAGTAAATAACACAGGTATACACATGCCATGGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 970 ACCTTATACCTAGCTCTCGAGGTGGGAGCATGATGATGATTTGGGCATCAATAT 1029
Db 361 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1038 ATCAAAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAA 1079
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLys 414

RESULT 5
AAW14733
ID AAW14733 standard; Protein; 423 AA.
XX
AC AAW14733;
XX
DT 25-MAR-2003 (updated)
DT 08-MAY-1997 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW plasma; plasminogen.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /note= "Signal peptide"
FT Protein 23..423
FT /note= "Mature PCPB"
FT Cleavage-site 114
FT /note= "Potential clip site for activation of PCPB
as a carboxypeptidase"
FT Active-site 181
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 183
FT /note= "Expected to be involved in substrate binding"
FT Active-site 184
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 256
FT /note= "Expected to be involved in substrate binding"
FT Binding-site 257
FT /note= "Expected to be involved in substrate binding"
FT Active-site 310
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 312
FT /note= "Expected to be involved in substrate binding"
FT Active-site 369
FT /note= "Expected to be involved in catalytic activity"
FT Binding-site 370
FT /note= "Expected to be involved in substrate
binding, thought to determine specificity
of PCPB as a carboxypeptidase B"

```

```

FT Binding-site 371
FT /note= "Expected to be involved in substrate binding"
FT Binding-site 394
FT /note= "Expected to be involved in substrate binding"
XX
XX US5593674-A.
XX 14-JAN-1997.
XX
XX 27-APR-1995; 95US-0430787.
XX
XX 01-FEB-1991; 91US-0649591.
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX 27-APR-1995; 95US-0430787.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1997-099413/09.
XX N-PSDB; AAT62846.
XX
XX Using human plasma carboxypeptidase B in blood coagulation - is
functionally related to carboxypeptidase A and pancreas
carboxypeptidase B
XX
XX Example 2; Column 37-42; 39pp; English.
XX
XX This sequence represents human plasma carboxypeptidase B (PCPB) which
has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
PCPB may be used therapeutically in haemostatic regulation. PCPB is
purified from human plasma or by transformed cell culture by
extraction using plasminogen bound to a solid phase.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 423 AA;

```

Alignment Scores:

```

Pred. No.: 1.29e-189 Length: 423
Score: 1835.00 Matches: 356
Percent Similarity: 86.23% Conservative: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 18 Gaps: 2

```

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAW14733 (1-423)

```

QY 1 ATGAAGCTTTGAGCCCTTGAGTCCCTTGACCCATTGTTCTTCTGTGAGCAGCATGTC 60
Db 1 MeCysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 61 TTCGCGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 121 GTTCTCAGAACTTACTACAAATAGATGAGATTGTTCTCTGGCAGCCGTAACAGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 181 CTTATTCTGAAGAAAAAACAAGTCCATTTTGTAAATGCATCTCATGTGCAATGTG 240
Db 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAlaAspVal 80
QY 241 AAAGCCCATTTAAATGTAGCGGAATTCATGTCAGTGTCTTGTGGCAGACGTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY 301 CTTATTCAACAGCAGATTTCACACGACACAGTACGAGCCCGAGCTCCGATCGTACTAT 360
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

```

```
QY 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGATTTATTAACCTGAGGCGAT 420
Db 121 GluGlnYrHisSerLeuAsnGluIleYrSerTrpIleGluPheIleThrGluArgHis 140
QY 421 CCTGATATGCTTACAAAATCCACATTCGATCCCTCATTTGAGAGTACCCACTCTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 481 TTAAGGTTTCTGAAAGAACAAACAGCCAAATATGCCATATGATGATGCTGTGGAATC 540
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 541 CATGCCAGAGATGATCTCTCTGCTTCTGCTGTTGCTGTTGCTATAGGCCAT 591
Db 181 HisAlaArgGlnTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 591 ----- 591
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 592 -----AATCGAATGTGGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 610 AACGTTCTTCTATGCGAACATCATTCGATCGAACACCTGATAGTACCACTTTGTC 669
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 670 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
QY 730 CTTTATCCTGAGTCAGAACAGAGTGAAGCAGTGGCTAGTCTTCTTGAAGAATAATC 789
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 790 AACGAGTTAAGCATATCATCAGCATCATTCATCTACTCCAGCATATAGTGTTCATAT 849
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 850 TCCTATACAGCACTAAAAGCAAGACCATGAGCACTGTCTCTAGTAGCCAGTGAAGCA 909
Db 321 SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 340
QY 910 GTTCGTCTATTGACAAAATAGTAAATAATACAGGTATATACATGSCCATGGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 970 ACCTTATACCTAGCTCTCTGAGGTGGGAGATGGATCTATGATTTGGGCATCAATAT 1029
Db 361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1038 ATCAAAACCCACCTGTAGAGAGCTTTTTCGCGTGTCTCTATAA 1079
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLys 414
RESULT 6
AAW92270
ID AAW92270 standard; Protein; 423 AA.
XX
AC AAW92270;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
XX polymorphism detection; thrombotic disease.
XX
```

```
OS Homo sapiens.
XX Key Location/Qualifiers
FT Peptide 1..22
FT Protein /note= "signal peptide"
FT /note= "mature PCPBthr147"
XX
PN WO9855645-A1.
XX 10-DEC-1998.
XX
XX 02-JUN-1998; 98WO-EP03244.
XX
XX 03-JUN-1997; 97US-0869057.
XX
XX (SCHD ) SCHERING AG.
XX
XX Morser MJ, Nagashima M;
XX
XX WPI; 1999-045800/04.
XX N-PSDB; AAV74302.
XX
XX Detecting new polymorphism of human plasma carboxypeptidase B -
XX comprises Alanine or Threonine at position 147 of protein by DNA or
XX protein analysis, useful to detect risk of thrombotic disease in
XX humans
XX
XX Disclosure; Page 25-26; 35pp; English.
XX
XX This sequence is the human plasma carboxypeptidase B (PCPB) mutant
XX hPCPBthr147. The invention relates to a method for determining the
XX presence of DNA or protein polymorphisms of PCPB in human subjects, which
XX comprises obtaining a prepared tissue or blood sample and determining the
XX presence of DNA coding for naturally occurring polymorphs of the protein
XX containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
XX respectively). Determination of the relative distribution of the PCPB
XX polymorphs in a patient's blood by genetic or protein analysis by the
XX methods is useful to determine the risk of thrombotic disease in humans.
XX Such assessments may be made by accumulating information concerning the
XX relative distribution of the different polymorphs within the general
XX population compared with populations known to be at risk and establishing
XX a PCPB polymorph profile for at-risk patients.
XX
SQ Sequence 423 AA;
Alignment Scores:
Pred. No.: 1,29e-189 Length: 423
Score: 1835.00 Matches: 356
Percent Similarity: 86.23% Conservative: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 20 Gaps: 2
US-09-980-881A-1_COPY_18_1097 (1-1080) x AAW92270 (1-423)
QY 1 ATGAAGCTTTCGACGCTTGACGCTTGACCCATTGTTCTCTCTGTGAGCAGCATGTC 60
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 61 TTCGCGTTCGAGAGTGGCCAAAGTTCAGCTGCTTCTTCTAGAACCTCTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 121 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCGCGTAAACGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 181 CTTATTGTGAAGAAAAACAAAGTCCATTTTGTAAATGTCATCTCATGTGCGAATGTG 240
Db 61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGTCAGTGTCTTGTGCGACAGCTGGAAGAT 300
```


Db 81 LysAlaHisLeuAenValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTATTCAACAGCAGATTTCCACGACACAGTACAGCCCGAGCCTCCGATCGTACTAT 360
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr 120
Qy 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGATTATTAACCTGAGAGCAT 420
Db 121 GluGlnIleHisSerLeuAenGluIleSerTrpIleGluPheIleThrGluArgHis 140
Qy 421 CCTGATATGCTTCAAAAATCCACATTCGATTCCTCATTTGAGAAGTACCCACTCTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Qy 481 TTAAGGTTTCTGAAAAGAACAAACAGCCAAAATATCCATATGATGATCTGTGGATC 540
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Qy 541 CATGCCAGAGATCGATCTCTCTGCTTCTGCTGTGGTTCATAGGCCAT----- 591
Db 181 HisAlaArgGluIleSerProAlaPheCysLeuIlePheIleGlyHisIleThrGln 200
Qy 591 ----- 591
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
Qy 592 -----AATCGAATGTGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
Qy 610 AACGGTTCTTCTATGCGAACAACTATTGTCATGCGACAGACCTGAATAGCACTTTGTC 669
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAenArgAsnPheAla 260
Qy 670 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTCGTCCGAAACCTACTGTGGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
Qy 730 CTTTATCTGAGTCAGAACCGAAGTGAAGCAGTGGCTAGTTCTTCTGAGAAGAAATATC 789
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 300
Qy 790 AACAGATTAAGCATACATCAGCATCATTCATCTCCAGCATATAGTTTCCATAT 849
Db 301 AsnGlnIleLysAlaIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
Qy 850 TCCTATACAGCAAGTAAAGCAAGACCATGAGCACTGTCTCTAGTAGCCAGTGAAGCA 909
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
Qy 910 GTTCGTGCTATTGACAAACTAGTAAATAACACAGGTATACATGCGCATGGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
Qy 970 ACCTTATACCTAGCTCTGAGGTCGAGCATGATGATCTATGTTGGCATCAATAT 1029
Db 361 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
Qy 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
Qy 1038 ATCAAAACCCACTGTAGAGAGCTTTTGGCCGCTCTCTCTAAA 1079
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLys 414

RESULT 7
AAR36273
ID AAR36273 standard; Protein; 423 AA.
XX
AC AAR36273;
XX

DT 24-AUG-1993 (first entry)
XX Human plasma carboxypeptidase B.
XX PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX Homo sapiens.
OS
PH Key Location/Qualifiers
FT Peptide /label= signal_peptide
FT 1..22
FT Cleavage-site /note= "putative site of cleavage by trypsin to
FT 114 activate PCPB as a carboxypeptidase"
FT
FT Active-site 181 /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 183 /note= "expected to be involved in substrate
FT binding"
FT Active-site 184 /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 256..257 /note= "expected to be involved in substrate
FT binding"
FT Active-site 310 /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 312 /note= "expected to be involved in substrate
FT binding"
FT Active-site 369 /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 370..371 /note= "expected to be involved in substrate
FT binding; Asp 370 is the residue which
FT determines specificity of the PCPB as
FT a carboxypeptidase B"
FT Binding-site 394 /note= "expected to be involved in substrate
FT binding"
XX US5206161-A.
XX 27-APR-1993.
XX 01-FEB-1991; 91US-0649591.
XX 01-FEB-1991; 91US-0649591.
XX (GETH) GENENTECH INC.
XX Drayna DT, Eaton DL;
XX WPI; 1993-151724/18.
XX N-PSDB; AAQ41001.
XX New human plasma carboxypeptidase B - used as haemostatic
XX regulator for clotting blood, partic. for treating blood clotting
XX disorders, e.g. haemophilia
XX Claim 2; Fig 4; 40pp; English.
XX This amino acid sequence was deduced from a clone isolated from a
XX human liver cDNA library. There is about 40% sequence identity
XX between the deduced preproPCPB amino acid sequence and pro-human
XX mast cell carboxypeptidase A and between preproPCPB and prepro-rat
XX carboxypeptidase B. Human PCPB has the same substrate binding sites
XX as, and shares 6 cysteine residues which form 3 intramolecular
XX disulphide bonds with, bovine and rat carboxypeptidase B. The
XX presence of the same amino acid (Asp at position 348 of the mature

CC protein) at the region in carboxypeptidases that determines
CC substrate specificity, suggests that PCPB represents a plasma-
CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC tPA of plasminogen to plasmin in the presence of fibrinogen.
xx
SQ Sequence 423 AA;
Alignment Scores:
Pred. No.: 4,49e-189 Length: 423
Score: 1830.00 Matches: 355
Percent Similarity: 85.99% Conservative: 1
Best Local Similarity: 85.75% Mismatches: 4
Query Match: 94.72% Indels: 55
DB: 14 Gaps: 2
US-09-980-881A-1_COPY_18_1097 (1-1080) x AAR36273 (1-423)
QY 1 ATGAAGCTTTGCGAGCTTGCAGTCTTGTAGTCCATTTCTCTCTCTGAGCAGCATGTC 60
DB 1 MetLysLeuCySerLeuAlaValLeuValProIleValLeuPheCyGluGlnHisVal 20
QY 61 TTCGGTTCAGAGTGCAGGCTTCTAGTCTCTCTCTAGAACCTCTAGSCAAGTTCAA 120
DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 121 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
DB 41 ValLeuGlnAsnLeuThrThrThrThrGluIleValLeuTrpGlnProValThrAlaasp 60
QY 181 CTTATTGTGAAGAAACCAAGTCCATTCTTTTGTAAATGCATCTGATGCGACATGTC 240
DB 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 241 AAAGCCCATTTAAATGTAGCGGAATTCACGAGTGTCTGTGTCGAGCAGCTGGAAGAT 300
DB 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaaspValGluasp 100
QY 301 CTTATTCAACAGCAGATTTCCACGACACAGTACGAGCCCCGAGCCTCCGATCGTACTAT 360
DB 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
QY 361 GAACAGTATCAGTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 420
DB 121 GluGlnIyrHisSerLeuAsnGluIleIleThrSerTrpIleGluPheIleThrGluArgHis 140
QY 421 CCTGATATGCTTCAAAAATCCCATTTGATGATCTCTCATTTGAGAAAGTACCCACTATGTT 480
DB 141 ProAspMetLeuThrLysLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 481 TTAAGGTTTCTGGAAGAAACAAACAGCCAAATAATGCCATATGATGATGATGGAATC 540
DB 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 541 CATGCCAGAGAAATGATCTCTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 591
DB 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 591 ----- 591
DB 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 592 -----AATCGAATGTGAGAAAG 609
DB 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 610 AACGGTCTTCTATGCGAACAATCATTTGATGATGGAACACCTGGAATAGCACTTTGTC 669
DB 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 670 TCCAAACACATGGTGTAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 729
DB 261 SerLysHisTrpCysGluGluGlyAlaCysSerSerCysSerGluThrTyrCysGly 280

QY 730 CTTTATCTCAGTCAGAACCAAGTGAAGGAGTGGCTAGTTCTTCTGAGAAATATC 789
DB 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 300
QY 790 AACAGATTAAAGCATACATCAGCATGCTTCTATCTCCAGCATATAGTGTTCATAT 849
DB 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 850 TCCTATACAGAAAGTAAAGCAAGACCATGAGAACTGTCTTAGTAGCCAGTGAAGCA 909
DB 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
QY 910 GTTCGCTGCTTATTCACAAACTAGTAAATAACAGGTATACATGCGCATGCTCGAA 969
DB 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 970 ACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCGCATCAATAT 1029
DB 361 ThrLeuTyrLeuAlaProGlyGlyGlyAspSerTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1030 TCCTT-----TAC 1037
DB 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1038 ATCAAAACCCACTGTAGAGAGCTTTTGGCGCTGCTCTTAAA 1079
DB 401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414
RESULT 8
AAB11458
ID AAB11458 standard; Protein; 338 AA.
XX
AC AAB11458;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; neurotropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WC2000066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Disclosure; Page 71-73; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent

CC diseases in which this occurs, such as Alzheimer's, senile dementia,
XX inherited cerebral hemorrhage, Down's syndrome, and head trauma.

SQ Sequence 338 AA;

Alignment Scores:

Pred. No.: 1.05e-183 Length: 338
Score: 1780.00 Matches: 335
Percent Similarity: 99.41% Conservative: 1
Best Local Similarity: 99.11% Mismatches: 2
Query Match: 92.13% Indels: 0
DB: 21 Gaps: 0

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAB11458 (1-338)

QY 67 TTCAGAGTGGCCAGTCTAGCTCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
QY 127 CAGAATCTTACTCAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGACCTTATT 186
Db 21 GlnAsnLeuThrThrThrGluLeuValLeuProGlnProValThrAlaAspLeuLe 40
QY 187 GTGAAGAAAAAACAAGTCCATTTTTTTTAAATGCATCTGATGTGCACAATGTGAAAGCC 246
Db 41 ValIysIysIysGlnValHisPhePheValAsnAlaSerAspValAspValIysAla 60
QY 247 CATTAAATGTGCGCGAATTCATGTCAGTGTCTTGTGCGCAGCGTGAAGATCTTATT 306
Db 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
QY 307 CAACAGCAGATTTCCACAGACAGTCAGCAGCCCGCCGACCTCCGATCGTACTATGACAG 366
Db 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerIleThrGluGln 100
QY 367 TATCAGTCACTAAATGAATCTATTTCTTGGATGAAATTTATACTGAGAGCGCATCTTGAT 426
Db 101 TyrHisSerLeuAsnGluIleThrSerTrpIleGluPheIleThrGluArgHisProAsp 120
QY 427 ATGCTTCAAAATTCACATGATGATCTCTCAATTTGAGAAAGTACCCATCTATGTTTAAAG 486
Db 121 MetLeuThrIysIleHisIleGlySerSerPheGluIysTyrProLeuTyrValLeuLys 140
QY 487 GTTCTCGAAAGAACAAACAGCCAAATGCCATATGATGATTCAGTGTGCAATCCATGCC 546
Db 141 ValSerGlyIysGluGlnThrAlaIysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
QY 547 AGAAGATGGATCTCTCTGCTTCTGCTTGTGGTTTCATAGGCCATATCGAATGTGGAGA 606
Db 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
QY 607 AAGAACGTTCTTCTATGGAAACATCATGTCATCGGAACAGACCTGGAATAGCACTTT 666
Db 181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
QY 667 GTCTCCAAACACGTGTGAGGAGGTGATCCAGTTCCTCATGCTCGGAACCTACTGT 726
Db 201 AlaSerIysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 220
QY 727 GGACTTTATCTCTAGTCAGAACACAGAGTGAAGCAGTGGCTAGTTTCTTGAGAGAAAT 786
Db 221 GlyLeuTyrProGluSerGluProGluValIysAlaValAlaSerPheLeuArgAsn 240
QY 787 ATCAACAGATTAAGCATCATCAGCATGATTCATCTCCAGCATATAGTGTTCCTCA 846
Db 241 IleAsnGlnIleIysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
QY 847 TATTCTTATACCAAGTAAAGCAAGACATGAGAACTGTCTCTAGTAGCCAGTGA 906
Db 261 TyrSerTyrThrArgSerIysSerLysAspHisGluLeuSerLeuValAlaSerGlu 280
QY 907 GCAGTTCGTCTATTGACAAAACCTAGTAAATAACACAGGTATACATGCGCCCTGCTCA 966
Db 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20

Db 281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
QY 967 GAAACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAA 1026
Db 301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGlyIleLys 320
QY 1027 TATTGCTTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAAA 1080
Db 321 TyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
RESULT 9
AAB11459
ID AAB11459 standard; Protein; 246 AA.
XX AAB11459;
AC AAB11459;
XX 01-MAR-2001 (first entry)
XX Human brain carboxypeptidase B protein SEQ ID NO 4.
XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX Homo sapiens.
OS WO200066717-A1.
XX PN 09-NOV-2000.
XX PD 01-MAY-2000; 2000WO-JP02878.
XX PF 30-APR-1999; 99JP-0125169.
XX PR (MATS/) MATSUMOTO A.
XX PA Matsumoto A;
XX PI WPI; 2000-687534/67.
XX DR Human brain carboxypeptidase B isolated from the hippocampus useful for
XX screening agents for the treatment of Alzheimer's and other brain
XX disorders -
XX PS Disclosure; Page 74-75; 84pp; Japanese.
XX This invention describes a novel protein with peptidase activity
XX against brain beta-amyloid precursor protein which has been isolated from
XX human hippocampus and which has cerebroprotective, antialzheimers,
XX nootropic, neuroprotective and hemostatic activity and which can be used
XX as a vaccine or for gene therapy. The protein, and compounds identified
XX by screening as promoters or inhibitors of its activity, are used to
XX regulate beta-amyloid accumulation in the brain and treat or prevent
XX diseases in which this occurs, such as Alzheimer's, senile dementia,
XX inherited cerebral hemorrhage, Down's syndrome, and head trauma.

SQ Sequence 246 AA;

Alignment Scores:
Pred. No.: 6.28e-134 Length: 246
Score: 1320.00 Matches: 243
Percent Similarity: 99.19% Conservative: 1
Best Local Similarity: 98.78% Mismatches: 2
Query Match: 68.32% Indels: 0
DB: 21 Gaps: 0

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAB11459 (1-246)

QY 343 GCCTCCGCATCGTACTATGACAGTATCACTCAATAAATGAAATCTATTCTCGATGAA 402
Db 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20

QY 403 TTTATACTGAGGCGATCTGATATGCTTACAAAAATCCACATTTGGATCCTCATTTGAG 462
Db 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
QY 463 AAGTACCCTCTATGTTTAAAGGTTTCTGGAAGAACAAACAGCCAAATGCGATA 522
Db 41 LysTyPProLeuTyValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
QY 523 TGGATTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTTCTGCTGTGCTTC 582
Db 61 TrpIleAspCysGlyIleHisAlaAArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
QY 583 ATAGGCCATATCGAATGTCGAGAGAAAGCGTTCTTTCTATGCGAAACATCATTCGATC 642
Db 81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnAsnHisCysIle 100
QY 643 GGAACAGACCTGAAATAGCACTTTGCTCCAAACACCTGGTGTGAGGAAGGTGCATCCAGT 702
Db 101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
QY 703 TCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCAAGAACAGAGTGAAGGCA 762
Db 121 SerSerCysSerGluThrTyCysGlyLeuTyProGluSerGluProGluValLysAla 140
QY 763 GTGGCTAGTTTCTTGAGAAATATCAACAGATTAAGCATACATCAGCATGCAATTC 822
Db 141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyIleSerMetHisSer 160
QY 823 TACTCCCGCATATAGTGTTCATATTCCTATACAGAGTAAAGCAAGACCATGAG 882
Db 161 TySerGlnHisIleValPheProTySerTyThrArgSerLysSerLysAspHisGlu 180
QY 883 GAACTGTCTCTAGTACGACAGTCAAGAGTTCGTGCTATTGACAAACCTAGTAAATATCC 942
Db 181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
QY 943 AGGTATACATGCCATGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGGACGAT 1002
Db 201 ArgTyThrHisGlyHisGlySerGluThrLeuTyLeuAlaProGlyGlyGlyAspAsp 220
QY 1003 TGGATCTATGATTGGGCATCAATATTCGTTTACATCAACCCACTGTGAGAGAGCTT 1062
Db 221 TrpIleTyAspLeuGlyIleLysTySerPheThrSerAsnProProValGluLysLeu 240
QY 1063 TTGCGCGCTGTCTCTAAAA 1080
Db 241 LeuProLeuSerLeuLys 246
RESULT 10
AAB58129
ID AAB58129 standard; Protein; 211 AA.
XX
AC AAB58129;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polypeptide sequence SEQ ID 467.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX

PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
DR WPI; 2000-587514/55.
DR N-PSDB; AAF18005.
XX
Lung cancer associated gene sequences, referred to as lung cancer
PT antigens useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -

Claim 11; Page 944-945; 1425pp; English.

XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterization of the polynucleotide and protein
CC sequences.

XX
SQ Sequence 211 AA;

Alignment Scores:
Pred. No.: 7.62e-82 Length: 211
Score: 839.00 Matches: 164
Percent Similarity: 82.59% Conservative: 2
Best Local Similarity: 81.59% Mismatches: 9
Query Match: 43.43% Indels: 27
DB: 21 Gaps: 3

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAB58129 (1-211)

QY 553 TGGATCTCTCTGCTTTCTGCTTGTGCTTCATA-----GGC 588
Db 3 TrpIleSer---MetLeuCysArgTrpLeuMetValMetAsnTySerTrpLys 21
QY 589 CATAATCGAATGTGGAGAAAGAACCGTTCTTCTATCGAACCAATCATTCATCGGAACA 648
Db 22 LysAsnArgMetTrpArgLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThr 41
QY 649 GACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATTCCTCA 708
Db 42 AspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSer 61
QY 709 TGCTCGGAACCTACTGTGAGCTTTTATCTGAGTCAGAACCAAGTGAAGCGGTGGCT 769
Db 62 CysSerGluThrTyCysGlyLeuTyProGluSerGluProGluValLysAlaValAla 81
QY 769 AGTTTCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCATCTCC 828
Db 82 SerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyIleSerMetHisSerTySer 101
QY 829 CAGCATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAAC 888
Db 102 GlnHisIleValPheProTySerTyThrArgSerLysSerLysAspHisGluGluLeu 121
QY 889 TCTTAGTAGCCAGTGAAGCAGTTCTGCTATTGACAAACTAGTAAATAACAGTAT 948
|||||

Db 316 TyAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335
 QY 913 CGTGCTATTGACAAAATACTAGTAAATACACAGTATACACATGCGCCATGCTCAGAAACC 972
 Db 336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354
 QY 973 TATACCTAGCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATATTCG 1032
 Db 355 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374
 QY 1033 TTTACATC----- 1040
 Db 375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394
 QY 1041 AAACCCACCTGTAGAAAGCTTTTGGCGGTGC 1073
 Db 395 ArgAlaThrCysGluGluThrPheLeuAlaIle 405
 RESULT 12
 AAR75131
 ID AAR75131 standard; Protein; 404 AA.
 AC AAR75131;
 XX
 DT 13-MAR-1996 (first entry)
 DE Porcine Tyr-His-Met Procarboxypeptidase B.
 XX
 KW Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
 KW human serum albumin; premating factor alpha; mating factor alpha;
 KW proCBB.
 XX
 OS Sus scrofa.
 PN W09514096-AL.
 XX
 PD 26-MAY-1995.
 PF 16-NOV-1994; 94WO-US13142.
 XX
 PR 16-NOV-1993; 93US-0153258.
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 PI Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
 PI Sterner JL, Zhang H;
 XX
 DR WPI; 1995-200386/26.
 DR N-PSDB; AAQ90600.
 XX
 PT DNA encoding porcine carboxypeptidase B - used for transforming
 PT host cells, partic. Pichia species, for prodn. of the enzyme
 XX
 PS Example 1; Page 16-17; 34pp; English.
 XX
 CC The porcine carboxypeptidase B coding sequence can be place in a
 CC bacterial or pref. Pichia yeast expression vector. The expression
 CC vector further comprises the signal peptide of either human serum
 CC albumin (designated pIGD23 - NRRL B-21029); premating factor alpha
 CC (designated pF489 - NRRL B-21028); mating factor alpha (designated
 CC pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
 CC (designated pLGD27 - NRRL B-21027). The method can be used for
 CC producing large amounts of porcine carboxypeptidase B and when
 CC produced in Pichia yeast, the protein does not need solubilisation
 CC or folding. The produced enzyme is then used for pref. cleaving
 CC basic residues from the carboxy terminus of proteins.
 XX
 SQ Sequence 404 AA;
 Alignment Scores:
 Pred. No.: 3,33e-60 Length: 404
 Score: 640.50 Matches: 136
 Percent Similarity: 54.29% Conservative: 60

Best Local Similarity: 37.67% Mismatches: 122
 Query Match: 33.15% Indels: 43
 DB: 16 Gaps: 6
 US-09-980-881A-1_COPY_18_1097 (1-1080) x AAR75131 (1-404)
 QY 67 TTCACAGATGGCCAAAGTTCTTAGCTGCTTCTTAGAACCTCTTAGGCAAGTTCAAGTTCTTA 126
 Db 10 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerLeuLeu 29
 QY 127 CAGAAATCTTACACATATGAGATTGTTCTCTGCAGCCGCTAACAGCTGACCTATT 186
 Db 30 HisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGlnIle 49
 QY 187 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTCATCTGCACATGTGAAAGCC 246
 Db 50 LysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGluAsp 69
 QY 247 CATTTAAATGTGAGCGGAATTCATGCAGTCTTCTCTGCAGACACGCTGGAGATCTTATT 306
 Db 70 PheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerValLeu 89
 QY 307 CAACAGCAGATTTCACACACACAGTCAGCCCGGAGCCTCCGCATCGTACTACTGAACAG 366
 Db 90 GluAlaGlnPheAspSerArgVal-----ArgThrThrGlyHisSerTyrGluLys 106
 QY 367 TATCACTCACTAAATGAAATCTATTCTGGATAGAAATTTATTAATGAGAGCATCTCTGAT 426
 Db 107 TyrAsnAsnTrpGluThrIleGluAlaTrpThrLysGlnValThrSerGluAsnProAsp 126
 QY 427 ATGCTTACAAAATCCACATTTGATCTCTCATTTGAGAAGTACCCTCTATGTTTAAAG 486
 Db 127 LeuIleSerArgThrAlaIleGlyThrThrPheLeuGlyAsnAsnIleTyrLeuLeuLys 146
 QY 487 GTTCTCGAAAACAAACACAGCCAAATGCCATATGGATTCACATGTGGAATCCATGCC 546
 Db 147 Val---GlyLysProGlyProAsnLysProAlaIlePheMetAspCysGlyPheHisAla 165
 QY 547 AGAGATGATCTCTCTCTCTTCTCTGTTGTTTCATA----- 585
 Db 166 ArgGluTrpIleSerHisAlaPheCysGlnTrpPheValArgGluAlaValLeuThrTyr 195
 QY 585 ----- 585
 Db 186 GlyTyrGluSerHisMetThrGluPheLeuAsnLysLeuAspPheTyrValLeuProVal 205
 QY 586 -----GGCCAT-----AATCGATGTGGAGAAAGACCGT 615
 Db 206 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThrArg 225
 QY 616 TCTTTCTATCGAAACAATCAATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
 Db 226 SerThrAsnAlaGlyThrThrCysIleGlyThrAspProAsnArgAsnPhe---AspAla 244
 QY 676 CACTGTGTGAGGAAGTGCATCCAGTTCTCATGTCTCGAAACCTACTGTGTGACTTTAT 735
 Db 245 GlyTrpCysThrThrGlyAlaSerThrAspProCysAspGluThrTyrCysGlySerAla 264
 QY 736 CCTGAGTCAGAACAGAGTGAAGGAGTGGCTAGTTCTTGTGAGAGAATATCAACCAG 795
 Db 265 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLeuSerSer 284
 QY 796 ATTAAGCATACATCAGCATGCATTCATCTACTCCAGCATATAGTGTTCATATTCCTAT 855
 Db 285 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 304
 QY 856 ACACGAAGTAAAGCAAGACCATAGGAACTGCTCTTAGTACCCAGTGAAGCAGTTGCT 915
 Db 305 AspTyrLysLeuProGluAsnAsnAlaGluLeuAsnAsnLeuAlaLysAlaValLys 324
 QY 916 GCTATTGACAAAACCTAGTAAATAACACAGGTATACATGCGCCATGGCTCAGAACTTA 975
 Db 325 GluLeu---AlaThrLeuTyrGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIle 343

QY	976	TACCTAGCTCCTGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAATAATTCGTTT	1035
Db	344	Tyr-Pro-Ala-alaglyGlySerAspAspTirAlaTyrAspGlnGlyIleLysTyrSer-Phe	363
QY	1036	ACA 1038	
Db	364	Thr 364	
RESULT 13			
AAW06172			
ID	AAW06172	standard; Protein; 415 AA.	
XX	AC		
XX	AAW06172;		
DT	12-FEB-1997	(first entry)	
XX			
DE	Human pancreatic carboxypeptidase B.		
XX			
KW	ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;		
KW	mustard-ribonucleotide; antibody directed enzyme prodrug therapy;		
KW	anti-neoplastic; prodrug; reverse polarity; ion pair interaction;		
KW	reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;		
KW	FG; F(ab') ₂ ; FcIB; leader; human carboxypeptidase B.		
XX			
OS	Homo sapiens.		
XX			
PH	Key	Location/Qualifiers	
FT	Peptide	1..13	
FT		/note= "partial enzyme secretion leader sequence;	
FT		(pre-sequence)"	
FT	Peptide	14..108	
FT		/label= pro-sequence	
FT	Protein	109..415	
FT		/label= mature_enzyme	
FT	Misc-difference	243	
FT		/note= "not observed in other published HPCB	
FT		sequences"	
FT	Active-site	361	
FT		/note= "Asp253 characteristic of B type specificity"	
XX			
PN	WO9620011-A1.		
XX			
PD	04-JUL-1996.		
XX			
XX	21-DEC-1995;	95WO-GB02991.	
PF			
XX			
PR	16-AUG-1995;	95GB-0016810.	
PR	23-DEC-1994;	94GB-0026192.	
XX			
PA	(ZENE) ZENECA LTD.		
XX			
PI	Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;		
PI	Hennam JF, Hennequin lFA, Marsham PR, Rabin BR, Slater AM;		
PI	Tarragona-Fiol A, Taylorson CJ;		
XX			
DR	WPI; 1996-321650/32.		
DR	N-PSDB; AAT42494.		
XX			
PT	Two component system for anti-tumour therapy - comprising targeting		
PT	moiety linked to mutated enzyme which can transform an		
PT	anti:neoplastic prodrug		
XX			
PS	Reference Example 14; Page 131-132; 182pp; English.		
XX			
CC	A two-component system for anti-tumour therapy comprises a targeting		
CC	moiety linked to a mutated enzyme which can transform an anti-neoplastic		
CC	prodrug. The system is based on antibody directed enzyme prodrug therapy		
CC	(ADEPT) using non-naturally occurring mutant forms of host enzymes,		
CC	pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).		
CC	Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can		
XX	be used. The present sequence is that of the native prepro-HCPB.		


```
QY 919 ATTGACAAACTAGTAAATACAGGTATACACATGGCCCATGGCTCAGAAACCTTATAC 978
Db :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257 Leu---LysValLeuPheGlyThrGluTyrThrTyrGlyProGlyAlaAlaThrIleTyr 275
QY 979 CTAGCTCCTGGAGTGGGACGATTGGATCTATGATTTGGGCATCAAATATTGTTTACA 1038
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 ProAlaAlaGlyGlySerAspTrrAlaTyrAspGlnGlyIleLysTyrAlaPheThr 295
QY 1039 TC-----AAACCC 1046
Db PheGluLeuArgAspLysGlyArgTyrGlyPheAlaLeuProGluSerGlnIleLysPro 315
QY 1047 ACCTGTAGAGAAGCTTTTGGCGGTGTC 1073
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 ThrCysGluGluThrMetIleAlaVal 324
```

Search completed: January 1, 2004, 18:52:58
Job time : 160.462 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:40:51 ; Search time 29.1067 Seconds
(without alignments)
3139.880 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagcttgagccttgc.....ttttgcgcgtgtctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0980881/runat_31122003_135750_24983/app_query.fasta_1.2958
-DB=Issued Patents AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0980881 @CGN 1.1 31 @runat_31122003_135750_24983 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1852.5	95.9	386	4	US-09-813-133A-2
2	1835	95.0	423	1	US-07-649-591B-3
3	1835	95.0	423	1	US-08-277-540-3
4	1835	95.0	423	1	US-08-430-787A-3
5	1835	95.0	423	2	US-08-869-057-2
6	1824	94.4	423	4	US-09-813-133A-4
7	640.5	33.2	404	1	US-08-696-139-2
8	639.5	33.1	415	2	US-08-860-882A-57
9	639.5	33.1	415	4	US-09-011-769A-39
10	638	33.0	417	1	US-07-649-591B-7
11	638	33.0	417	1	US-08-277-540-7
12	638	33.0	417	1	US-08-430-787A-7

13	635	32.9	417	1	US-07-649-591B-6	Sequence 6, Appli
14	635	32.9	417	1	US-08-277-540-6	Sequence 6, Appli
15	635	32.9	417	1	US-08-430-787A-6	Sequence 6, Appli
16	632	32.7	424	4	US-09-011-769A-56	Sequence 56, Appl
17	623	32.2	716	3	US-09-171-945-125	Sequence 125, App
18	613.5	31.8	437	4	US-09-675-305-10	Sequence 10, Appl
19	604	31.3	307	2	US-08-782-760-6	Sequence 6, Appli
20	604	31.3	307	5	PCT-US96-00995-6	Sequence 6, Appli
21	603.5	31.2	336	1	US-07-649-591B-4	Sequence 4, Appli
22	603.5	31.2	336	1	US-08-277-540-4	Sequence 4, Appli
23	603.5	31.2	336	1	US-08-430-787A-4	Sequence 4, Appli
24	576	29.8	306	1	US-08-696-139-4	Sequence 4, Appli
25	561	29.0	329	4	US-09-011-769A-51	Sequence 51, Appl
26	561	29.0	329	4	US-09-011-769A-47	Sequence 47, Appl
27	560	29.0	613	4	US-09-171-945-113	Sequence 113, App
28	554	28.7	349	4	US-09-011-769A-60	Sequence 60, Appl
29	553	28.6	349	4	US-09-011-769A-64	Sequence 64, Appl
30	553	28.6	417	1	US-07-649-591B-8	Sequence 8, Appli
31	553	28.6	417	1	US-08-277-540-8	Sequence 8, Appli
32	553	28.6	417	1	US-08-430-787A-8	Sequence 8, Appli
33	546	28.3	339	4	US-09-710-099-8	Sequence 8, Appli
34	528.5	27.4	350	4	US-09-675-305-12	Sequence 12, Appl
35	528.5	27.4	419	1	US-07-649-591B-5	Sequence 5, Appli
36	528.5	27.4	419	1	US-08-277-540-5	Sequence 5, Appli
37	528.5	27.4	419	1	US-08-430-787A-5	Sequence 5, Appli
38	519.5	26.9	436	4	US-09-710-099-6	Sequence 6, Appli
39	513	26.6	417	3	US-08-640-906-4	Sequence 4, Appli
40	513	26.6	417	4	US-09-395-936-4	Sequence 4, Appli
41	512	26.5	417	3	US-08-640-906-18	Sequence 18, Appl
42	512	26.5	417	4	US-09-395-936-18	Sequence 18, Appl
43	487.5	25.2	419	3	US-08-640-906-2	Sequence 2, Appli
44	487.5	25.2	419	4	US-09-395-936-2	Sequence 2, Appli
45	482.5	25.0	419	3	US-08-640-906-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Alignment Scores:
Pred. No.: 9.6e-209 Length: 386
Score: 1852.50 Matches: 354
Percent Similarity: 94.16% Conservative: 1
Best Local Similarity: 93.90% Mismatches: 5
Query Match: 95.89% Indels: 18
DB: 4 Gaps: 1

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-813-133A-2 (1-386)

Qy	1	ATGAAGCTTTCAGCCTTGAGTCCTTGTACCCATTGCTCTTCTGTGAGCAGCATGTC	60
Db	1	MetLysLeuCysSerLeuValLeuValProIleValLeuPheCysGluGlnHieVal	20
Qy	61	TTGCGGTTCCAGAGTGCCCAAGTTCTAGCTGCTTCTTAGACCTCTAGCAAGTTCAA	120

Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGATCTTACTACACATATGAGATGTTCTCTGGCAGCGGTAAACAGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrTyGluValLeuTrpGlnProValThrAlaAsp 60
Qy 181 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 240
Db 61 LeuIleVallylsLysGlnValHisPhePheValAsnAlaSerAspValAspVal 80
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTCTCTGTCGGCAGACGTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGATTTCCACGACACAGTCAAGCCCGAGCCCTCCGATCGTACTAT 360
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyTyr 120
Qy 361 GAACAGTATCACTCAATAATGAATCTATTCTTGGATAGAATTATAACTGAGAGCAT 420
Db 121 GluGlnTyHisSerLeuAsnGluIleTySerTrpIleGluPheIleThrGluArgHis 140
Qy 421 CTTGATATGCTTCAAAAAATCCCATTTGATTCCTATTGAGAAAGTACCCACTCTATGT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeuTyVal 160
Qy 481 TTAAGGTTCTCGAAGAACACACAGCCAAATGCCATATGATGATGCTGGAAATC 540
Db 161 LeuLysValSerGlyLysGluGlnAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Qy 541 CATGCCAGAGATGGATCTCTCTGCTTCTGCTGTGTTGTTGTTGTTGTTGTTGTTGTTG 600
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
Qy 601 TGGAGAAAGAACCGTTCTTCTATCGAACAATCATTCATCGGAACAGACCTGAATAGC 660
Db 201 TrpArgLysAsnArgSerPheTyAlaAsnHisCysIleGlyThrAspLeuAsnArg 220
Qy 661 AACTTTCTCCAAACACTGGTGTGAGAAAGTGCATCCAGTTCCTCATCTCGGAATC 720
Db 221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThr 240
Qy 721 TACTGTGAGATTTATCTGAGTCAAGACCAAGAGTGAAGCGAGTGTCTTCTTCTGAGA 780
Db 241 TyrCysGlyLeuTyProGluSerGluProGluVallysalavalAlaSerPheLeuArg 260
Qy 781 AGAATATCAACAGATTAAGCATACATCAGCATGCAATTCATCTACTCCAGCATATAGT 840
Db 261 ArgAsnIleAsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleVal 280
Qy 841 TTTCATATTCCTATACAGAGTAAAGCAAGACCATGAGGACTGTCTCTAGTAGCC 900
Db 281 PheProTySerTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
Qy 901 AGTGAAGCAGTTCGTGCTATTGACAAACATAGTAAAAATACAGGTATACACATGGCCAT 960
Db 301 SerGluAlaValAlaGluAlaGluLysIleSerLysAsnThrArgTyThrHisGlyHis 320
Qy 961 GGCTCAGAAACCTTATACCTAGTCTCTGAGAGTGGGACGATGGATCTATGATTTGGGC 1020
Db 321 GlySerGluThrLeuTyLeuAlaProGlyGlyAspAspTrpIleTyAspLeuGly 340
Qy 1021 ATCAAAATTCGTT----- 1034
Db 341 IleLysTySerPheThrIleGluLeuArgAspThrGlyThrTyGlyPheLeuLeuPro 360
Qy 1035 -----TACATCAAAACCCCTGTAGAGAGCTTTTCCGCTGTCTCTAAA 1079
Db 361 GluArgTyIleLysProThrCysArgGluAlaPheAlaAlaValSerLys 377

RESULT 2

US-07-649-591B-3

; Sequence 3, Application US/07649591B

; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Baton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Alignment Scores:
Pred. No.: 1.15e-206 Length: 423
Score: 1835.00 Matches: 356
Percent Similarity: 86.23% Conservative: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-07-649-591B-3 (1-423)

Qy 1 ATCAAGCTTTGACAGCTTGTGAGTCCCTGTACCCATTTGTTCTTCTGTGAGCAGCATGTC 60
Db 1 MetCysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Qy 61 TTGCGCTTCACAGAGTGCACCAAGTTCCTAGCTGCTCTTCTAGAACCTCTTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGATCTTACTACACATATGAGATTTGTTCTCTGCGAGCGGTAAACAGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrTyGluIleValLeuTrpGlnProValThrAlaAsp 60
Qy 181 CTTATTGTGAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 240
Db 61 LeuIleVallylsLysGlnValHisPhePheValAsnAlaSerAspValAspVal 80
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTCTCTTGTGCGACAGCTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAAGCCCGAGCCCTCCGATCGTACTAT 360
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyTyr 120


```
Qy 421 CCTGATATGCTTACAAAAATCCACATTTGGATCTCTATTGAGAAATACCCACTTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Qy 481 TTTAAAGTTTCTGGAAGAACAAACAGCCAAATAATGCCATATGATGATGATGGAATC 540
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Qy 541 CATGCCAGAGAAATGATCTCTCTGCTTTCTGCTTGGTTCATAGCCCAT----- 591
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
Qy 591 ----- 591
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
Qy 592 -----AATCGAATGTGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
Qy 610 AACCGTTCTTCTATGGAACAAATCATTTGATCGGAACAGACACTGAATAGCACTTTGTC 669
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
Qy 670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
Qy 730 CTTTATCCTGAGTCAGAACAGAGTGAAGCGAGTGTCTTCTTGTGAGAAGAAATATC 789
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAlaGAsnIle 300
Qy 790 AACGAGATTAAAGATACATCAGCATCATCTACTTCCAGCATATAGTGTTCATAT 849
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
Qy 850 TCCTATACAGCAAGTAAAGCAAGACCATGAGAACTGTCTTAGTAGCAGTGAAGCA 909
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAla 340
Qy 910 GTTCGTCTATTGAAAACTAGTAAATAATACCAAGGTATACATGCGCCATGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
Qy 970 ACCTTATACCTAGTCTCTGAGGTGGGAGATGATGATGATGATGATGATGATGATGAT 1029
Db 361 ThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGlyIleLysTyr 380
Qy 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
Qy 1038 ATCAAAACCCACCTGTAGAGAGCTTTTTCGCGCTGTCTCTAAA 1079
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLys 414
```

RESULT 4

```
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-430-787A-3
```

```
Alignment Scores:
Pred. No.: 1,15e-206 Length: 423
Score: 1835.00 Matches: 356
Percent Similarity: 86.23% Conservative: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 2 Gaps: 2
```

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-430-787A-3 (1-423)

```
Qy 1 ATCAAGCTTTCAGACCTTCAGCTTCCTGTACCATTTGTTCTCTCTGTGAGCAGCATGTC 60
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Qy 61 TTCGCGTTTCAGAGTGGCCAAAGTTCTAGCTCTTCTAGAACCTCTTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGATCTTACTACACATATGAGATTTGTTCTCTGGCAGCCGGTAAACAGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
Qy 181 CTTATTGTGAGAAAAACAAGTCCATTTTGTAAATGCATCTCATGTCGCAATGTC 240
Db 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Qy 241 AAAGCCCATTTAAATGTCGAGCGGAAATCCATGTCAGTGTCTTGTGCGACAGCTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGATTTCCAAACAGACACAGTCAGCCCGCCCGACCTCCGCTACTAT 360
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Qy 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAATTTTAACTAGAGAGCAT 420
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
Qy 421 CTTGATATGCTTACAAAAATCCACATTTGGATCTCTATTGAGAAATACCCACTTATGTT 480
```



```
QY 541 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGGTTTCATAGGCCAT-----591
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 591 -----591
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 592 -----AATCGAATGTGGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 610 AACCGTTCTTCTATCGGAACAATTCATTGCTCGAAGACAGACCTGAATAGCACTTTGTC 669
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 670 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTTCCTCATGCTCGGAACCTACTGTGCGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
QY 730 CTTTATCTGAGTCAGAACCGAAGTGAAGCAGTGGCTAGTTCTTCTGAGAAGAAATATC 789
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 790 AACAGATTAAAGCATACATCAGCATGCTTCTACTCCAGCATATAGTCTTCCATAT 849
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 850 TCCTATACACGAAGTAAAGCAAGACCATGAGAACTGCTCTAGTAGCAGTGAAGCA 909
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
QY 910 GTTCGTGCTATTGACAAACTAGTAAATAACAGGTATACATGCGCATGGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 970 ACCTTATACCTAGCTCTGGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATAT 1029
Db 361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1038 ATCAAAACCCCTCTAGAGAGCTTTTGGCGCTCTCTCTAAA 1079
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLys 414
```

RESULT 6

```
US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-4

Alignment Scores:
Pred. No.: 2,25e-205 Length: 423
Score: 1824.00 Matches: 354
Percent Similarity: 85.75% Conservative: 1
```

```
Best Local Similarity: 85.51% Mismatches: 5
Query Match: 94.41% Indels: 55
DB: 4 Gaps: 2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-813-133A-4 (1-423)

QY 1 ATGAAGCTTTGCGAGCCTTGAGTCTTGTATCCCATCTTCTCTCTCTGAGCAGCATGTC 60
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 61 TTCGCGTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 121 GTTCTACAGAAATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 181 CTTATTGTGAAGAAAAACAAGTCCATTCTTTTGTAAATGCATCTGATGTCGACAAATGTG 240
Db 61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGGCAGACGTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
QY 301 CTTATTCAACAGCAGATTTCCAAAGCACACAGTACGCCCCCGAGCCTCCGATGCTACTAT 360
Db 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr 120
QY 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGATTATTAACGTAGAGGAT 420
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 421 CCTGATATGCTTACAAAATCCACATTGGATCTCTCACTTTGAGAAGTACCCACTCTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 481 TTTAAGGTTTCTGAAAGAAACAAACAGCCAAAATGCCATATGCGATTGATGCTGGAAATC 540
Db 161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 541 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGGTTTCATAGGCCAT-----591
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 591 -----591
Db 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 592 -----AATCGAATGTGGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 610 AACCGTTCTTCTATCGGAACAATTCATTGCTCGAAGACAGACCTGAATAGCACTTTGTC 669
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 670 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTTCCTCATGCTCGGAACCTACTGTGCGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
QY 730 CTTTATCTGAGTCAGAACCGAAGTGAAGCAGTGGCTAGTTCTTCTGAGAAGAAATATC 789
Db 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 790 AACAGATTAAAGCATACATCAGCATGCTTCTACTCCAGCATATAGTCTTCCATAT 849
Db 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 850 TCCTATACACGAAGTAAAGCAAGACCATGAGAACTGCTCTAGTAGCAGTGAAGCA 909
Db 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
```



```
QY 1036 ACA 1038
Db 364 Thr 364

RESULT 8
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-860-882A-57

Alignment Scores:
Pred. No.: 3,35e-66 Length: 415
Score: 639.50 Matches: 140
Percent Similarity: 51.72% Conservative: 71
Best Local Similarity: 34.31% Mismatches: 136
Query Match: 33.10% Indels: 62
DB: 2 Gaps: 8

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-860-882A-57 (1-415)
QY 16 CTTGACGTCCTTGACCATGTTCTCTCTGTGAGCAGCATGTCCTTCGCGG---TTCCAG 72
Db 2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21
QY 73 AGTGCCCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAA 132
Db 22 GlyGlyValPheArgValAsnValGluAsnHisIleAsnIleArgGlu 41
QY 133 CTTACTACAACATATGAGATTGTTCTCTGCGCAGCGCGGTAAACAGCTGACCTTATTGTGAAG 192

42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61
193 AAAAAACAAGTCCATTTTGTAAATGATCTGATGTCGACAAATGTAAAGCCCATTTA 252
62 HisSerThrValAspPheArgValLysAlaGluAsnValThrValGluAsnValLeu 81
253 AATGTGAGCGGAATTCATCGACGTGCTGCTGCGCAGAGCTGGAAGATCTTATTCAACAG 312
82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101
313 CAGATTTTCCAACGACAGCAGTCAGCCCGAGCCCTCGCATCTGCTACTGATCAACAGTATC 372
102 GlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118
373 TCACTAAATGAATCTATTCTTGGATAGAAATTTATTAAGTGAAGGATCTCGATATGCTT 432
119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138
433 ACAAAATCCACATTCGATTCCTCATTTGAGAGTACCACCTCTATGTTTAAAGGTTTCT 492
139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157
493 GGAAGAAACAACAGACCAAAATGCATATGATGATTCACCTGCTGGAATCCATGCCAGAGAA 552
158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177
553 TGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 585
178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197
585 -----
198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217
586 -----GGCCAT-----AATCGAATGTGAGAGAAAGAACCGTCTCTTTC 621
218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237
622 TATCGAACAATCATTTGCAATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256
682 TGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGACCTTTATTCCTGAG 741
257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276
742 TCAGAACCAAGTGAAGCAGTGGCTAGTTTCTTGAGAGAAATATCAACAGATATAA 801
277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296
802 GCATACATCAGATGATTCATATCTCCAGCATATAGTGTGTTCCATATTCCTATACACGA 861
297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316
862 AGTAAAGCAACACCATCAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
317 LysLeuGlyGluAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336
922 GACAAACTAGTAAATAACACAGTATACATGCGCATCGGCGCATCGGCGCATCTATACCTA 981
337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355
982 GCTCCTGGAGGTGGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1040
356 AlaAlaGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375
1041 -----AAACCCACC 1049
376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395
1050 TGTAGAGAACTTTTTCGCGCTGTC 1073
1073 -----
```

Db 396 CysGluGluThrPheLeuAlaIle 403

RESULT 9

US-09-011-769A-39

; Sequence 39, Application US/09011769A

; Patent No. 6436691

GENERAL INFORMATION:

APPLICANT: SLATER, Anthony M.

BLAKEY, David C.

DAVIES, David H.

HENNAM, John F.

HENNEQUIN, Laurent F.A.

MARSHAM, Peter R.

DOWELL, Robert I.

TITLE OF INVENTION: Chemical Compounds

NUMBER OF SEQUENCES: 87

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Pillsbury Madison & Sutro, LLP

STREET: 1100 New York Ave., N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 Mb disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/011,769A

FILING DATE: 13-Feb-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB96/01975

FILING DATE: 13-AUG-1996

APPLICATION NUMBER: GB 9612295.7

FILING DATE: 12-JUN-1996

APPLICATION NUMBER: GB 9611019.2

FILING DATE: 25-MAY-1996

APPLICATION NUMBER: GB 9516810.0

FILING DATE: 16-AUG-1995

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:

LENGTH: 415 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 39:

US-09-011-769A-39

Alignment Scores:

Pred. No.: 3.35e-66 Length: 415

Score: 639.50 Matches: 140

Percent Similarity: 51.72% Conservative: 71

Best Local Similarity: 34.31% Mismatches: 136

Query Match: 33.10% Indels: 62

DB: 4 Gaps: 8

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-011-769A-39 (1-415)

Qy 16 CTTGACGCTCTTGACCATGTTCTCTCTGTGAGCAGCATGCTTCGCG---TTCCAG 72

Db 2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisGlyGlyGlyPheGlu 21

Qy 73 AGTGCCCAAGTTCTAGCTGCTCTCTAGACACCTCTAGGACAGTCAAGTCTACAGAT 132

Db 22 GlyGlyValPheArgValAsnValGluAspGluAsnHisIleAsnIleArgGlu 41

Qy 133 CTTACTACACATATAGATTGTTCTCTGGCAGCGCGTACAGCTGACCTTATTGTGAAG 192

Db 42 LeuAlaSerThrThrGlnIleAspPheTrpIysProAspSerValThrGlnIleLysPro 61

Qy 193 AAAAAACAAGTCCATTTTGTAAATGCACTGTGATGTCGACAAATGTGAAGCCCATTTA 252

Db 62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy 253 AATGTGAGCGGAATTCATCGAGTGTCTGTCGACAGCTGGAAGATCTTATTCAACAG 312

Db 82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValGluAla 101

Qy 313 CAGATTTTCAACGACACAGTCCGCGCCGCGCTCTGCTACTGATGAACGATATCAC 372

Db 102 GlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGlyLysTyrAsn 118

Qy 373 TCACTAAATGAATCTATTCTTGGATAGAATTTATACTGAGAGGCACTCTGATATGCTT 432

Db 119 LysTrpGluThrIleGluAlaTrpThrGlnValAlaThrGluAsnProAlaLeuIle 138

Qy 433 ACAAAATCCACATGTCATCTTGAAGTACCCACTCTATGTTTAAAGGTTTCT 492

Db 139 SerArgSerValIleGlyThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy 493 GGAAGAACAAACAGCCAAATGCCATATGGAATGACTGTGGAATCCATGCCAGAGAA 552

Db 158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy 553 TGAATCTCTCTGCTTCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 585

Db 178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy 585 ----- 585

Db 198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy 586 -----GGCCAT-----AATCGAATGTGAGAAAGAACCGTTCTTCTTC 621

Db 218 IleAspGlyTyrIleTyrThrThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy 622 TATGCGAACAAATCATTCATCGGACAGACCTGAATAGCAACTTGTCTCAACACTGG 681

Db 238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy 682 TGTGAGGAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGACCTTATCTCTGAG 741

Db 257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy 742 TCAGAACCAAGTGAAGCAGTGGTGTCTTCTTGAGAAAGAAATATCAACCAATTAAA 801

Db 277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerLys 296

Qy 802 GCATACATCAGCATGCTATCATCTCCAGCATATAGTGTGTTTCCATATCTCTATACAG 861

Db 297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy 862 AGTAAAGCAACACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921

Db 317 LysLeuGlyGluAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy 922 GACAAACTAGTAAATAATACAGGTATACATGCGCCATCGCTCAGAACTTATACCTA 981

Db 337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy 982 GCTCTCGGAGGTGGGACGATTCATGATTTGGGCATCAAAATATCTGTTTACATC- 1040

Db 356 AlaAlaGlyGlySerAspSerTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy 1041 -----AAACCCACC 1049

Db 376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy 1050 TGTAGAGAGCTTTTCCGCTGTC 1073

Db 396 CysGluGluThrPheLeuAlaIle 403

RESULT 10

CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: patin (Genentech)
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/277,540
 FILING DATE: 19-JUL-1994

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/167727
 FILING DATE: 15-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/959944
 FILING DATE: 14-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/649591
 FILING DATE: 01-FEB-91

ATTORNEY/AGENT INFORMATION:

NAME: Haesak, Janet E.

REGISTRATION NUMBER: 28,616

REFERENCE/DOCKET NUMBER: 689D1C1D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1896

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 417 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-277-540-7

Alignment Scores:

Pred. No.: 5,04e-66 Length: 417
 Score: 638.00 Matches: 147
 Percent Similarity: 51.23% Conservative: 62
 Best Local Similarity: 36.03% Mismatches: 135
 Query Match: 33.02% Indels: 65
 DB: 1 Gaps: 9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-277-540-7 (1-417)

```

Qy 16 CTTGCACTCTTGACCATGTTCTCTCTGTGAGCAGCATGTCTTCGGCTTCCAGAGT 75
Db 7 MetAlaValIleTyrThrLeuAlaIleAlaProValHis-----PheAspArg 23
Qy 76 GGCCAAGTTCTAGCTGCTCTCTCCAGAACCTCTAGGCAAGTTCAAGTTCTACAGAACTTT 135
Db 24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43
Qy 136 ACTACACATATGAGATTGTTCTCTGGCAGCCGCTAACAGCTGACCTTATTGTGAAGAAA 195
Db 44 ThrGlnSerIleGluLeuAspPheTyrProAspAlaIleHisAspIleAlaValAsn 63
Qy 196 AAACAAGTCCATTTTGTAAATGCTGATCTGCACATGTGAAGCCCATTTAAAT 255
Db 64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83
Qy 256 GTGAGCGGAATTCATGCTGCTCTGTCGCACAGCTGGAAGATCTTATTCAACACAGC 315
Db 84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluIleGluLysGln 103
Qy 316 ATTTCC---AACGACACAGTCAGCCCGAGCCCTCCGCTACTATGACAGTATCAC 372
Db 104 PheAspValLysAspGluIleAlaGlyArgHisSer-----TyrAlaLysTyrAsn 120
Qy 373 TCACATAATGAATCTATTCTTGATAGATTATTAATCACTGAGAGCATCTCTGATGCTT 432

```

```

Db 121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140
Qy 433 ACAAAATCCACATTTGATCTCTCATTTGAGAGTACCCTACTATGTTTAAAGTTTCT 492
Db 141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159
Qy 493 GGAAGAACAACAGCCAAATATGATATGATGATGATGATGATGATGATGATGATGATGAT 552
Db 160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179
Qy 553 TGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 591
Db 180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199
Qy 592 AAT----- 594
Db 200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219
Qy 595 -----CGAATGTGGAGAGAAACCGTTCTTTC 621
Db 220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239
Qy 622 TATGCGAACAATCATTTGATCGGACGACCTGATAGCACTTGTCTCCAAACACTGG 681
Db 240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 259
Qy 682 TGTGAGGAGGTGCATCCAGTCTCTCATGTCGCGAAACCTACTGTCGACTTTTCTCTGAG 741
Db 259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278
Qy 742 TCAGAACCAAGTGAAGGAGTGGTGTCTTCTTGAGAGAAATATCAACACGATTTAA 801
Db 279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
Qy 802 GCATACATCAGCATGCTACTCTCCAGCATATAGTGTCTTCATATCTCTATACAGA 861
Db 299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318
Qy 862 AGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTCTATT 921
Db 319 LysLeuProProAsnHisGlnAspLeuLysValAlaArgIleAlaThrAspAlaLeu 338
Qy 922 GACAAACTAGTAAATAACAGGTATACAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT 981
Db 339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357
Qy 982 GCTCTCGAGGTGGGACGATTCGATCTATGATTTGGGCATCAATATTCGTTTAC---- 1037
Db 358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377
Qy 1038 -----ATCAAAACCCACC 1049
Db 378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397
Qy 1050 TGTAGAAGCTTTTCCCGCTGTC 1073
Db 398 CysLysGluThrMetLeuSerVal 405

```

RESULT 12

US-08-430-787A-7
 ; Sequence 7, Application US/08430787A
 ; Patent No. 5593674
 ; GENERAL INFORMATION:
 ; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
 ; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSSEE: Genentech, Inc.
 ; STREET: 460 Point San Bruno Blvd
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-7

Alignment Scores:

Pred. No.:	5.04e-66	Length:	417
Score:	638.00	Matches:	147
Percent Similarity:	51.23%	Conservative:	62
Best Local Similarity:	36.03%	Mismatches:	135
Query Match:	33.02%	Indels:	65
DB:	1	Gaps:	9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-430-787A-7 (1-417)

Qy	16	CTTCAGTCTCTGACCATGTTCTCTCTGAGCAGCATGTTCTCGGCTTCACAGT	75
Db	7	MetAlaValIleTyrThrLeuAlaIleAlaProValHis-----PheAspArg	23
Qy	76	GGCCAACTTCTAGCTCTCTCTAGAACCTCTAGGCAAGTTCAGTCTCTACAAATCTT	135
Db	24	GlulysValPheArgValLysLeuGlnAsnGlulysHisAlaSerValLeuLysAsnLeu	43
Qy	136	ACTACACATATGAGATTGTTCTCTGGCAGCGCGTAACAGCTGACCTTATTGTGAAGAA	195
Db	44	ThrGlnSerIleGluLeuAspPheTyrProAspAlaIleHisAspIleAlaValAsn	63
Qy	196	AAACAGTCCATTTTGTGTAATGTCATCTGACATGTCGACATGTGAAGCCATTTAAAT	255
Db	64	MetThrValAspPheArgValSerGlulysGlnThrIleGlnSerThrLeuGlu	83
Qy	256	GTGAGCGGAATTCATGTCAGTCTCTGTCGCACACGTCGGAAGATCTTATCAACACAG	315
Db	84	GlnHisIleHisTyrGluIleIleHisAspLeuGlnGluIleGluLysGln	103
Qy	316	ATTTC-----AACGACACAGTCAGCGCCCGACCTCCGATCGTACTATGAACATGAC	372
Db	104	PheAspValLysAspGluIleAlaGlyArgHisSer-----TyrAlaLysTyrAsn	120
Qy	373	TCACATAAGAAATCTATTTTGGATGAATTTATTAAGTACGAGCGCATCTGTATGCTT	432
Db	121	AspTrpAspIleValSerTrpThrGlulysMetLeuGlulysHisProGluMetVal	140

Qy	433	ACAAAATCCACATTGGATCTCTATTTGAGAAGTACCCTACTATGTTTAAAGTTTCT	492
Db	141	SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle---	159
Qy	493	GGAAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAA	552
Db	160	GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu	179
Qy	553	TGATCTCTCTGCTTCTCTCTGTTGTTGTTCTATA-----GGCCAT	591
Db	180	TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys	199
Qy	592	AAT-----	594
Db	200	AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn	219
Qy	595	-----CGAATGTGGAGAAAGAACCGTCTTTTC	621
Db	220	ValAspGlyTyrIleTyrSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg	239
Qy	622	TATGCCGACAAATCATTCATCGCAACAGACTCAATAGCAACTTGTCTCCAAACACTGG	681
Db	240	AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp	258
Qy	682	TGTGAGGAAGTGCATCCAGTCTCTCATGCTCGGAAACCTACTGTGACTTTTATCCCTGAG	741
Db	259	AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu	278
Qy	742	TCAGAACCAAGTGAAGCAGTGGCTAGTCTTCTTGAGAAGAAATATCAACCAATATAA	801
Db	279	SerGlulysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys	298
Qy	802	GCATACATCAGCATGCTATCTACTCCAGCATATAGTGTTCATATTCCTATACACGA	861
Db	299	AlaTyrIleThrPheHisSerTyrSerGlnMetLeuIleProTyrGlyTyrThrPhe	318
Qy	862	AGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT	921
Db	319	LysLeuProProAsnHisGlnAspLeuLysValAlaArgIleAlaThrAspAlaLeu	338
Qy	922	GACAAAATAGTAAATAATACAGGTATACATACATGCGCATGGCTCAGAAACCTTATACCTA	981
Db	339	---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys	357
Qy	982	GCTCTCGAGTGGGACGATTCGATCTATGATTTGGCATCAATATTCGTTTAC----	1037
Db	358	ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe	377
Qy	1038	-----ATCAAAACCCACC	1049
Db	378	GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr	397
Qy	1050	TGTAGAGAGCTTTTGGCGCTGTC	1073
Db	398	CysLysGluThrMetLeuSerVal	405

RESULT 13
US-07-649-591B-6
Sequence 6, Application US/07649591B
Patent No. 5206161
GENERAL INFORMATION:
APPLICANT: Dennis Drayna and Daniel Eaton
TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasek, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/952-9881
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-6

```

```

Alignment Scores:
Pred. No.: 1,14e-65 Length: 417
Score: 635.00 Matches: 139
Percent Similarity: 52.31% Conservative: 76
Best Local Similarity: 33.82% Mismatches: 132
Query Match: 32.87% Indels: 65
DB: 1 Gaps: 9

```

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-07-649-591B-6 (1-417)

```

Qy 16 CTTGCACTCTTGACCAATGTTCTCTCTGTCGAGCAGCATGCTTCGCG-----66
Db 1 MetArgLeuIleuProValGlyLeuIleAlaThrLeuAlaIleAlaProValArg 20
Qy 67 TTCAGAGTGGCCAAAGTTTCTAGTCTCTCTAGAACCTCTAGGCAAGTTCAGTTCTA 126
Db 21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIle 40
Qy 127 CAGAACTTACTACAAATAGAAATGTTCTCTGCGAGCGGTAACAGCTGACCTTATT 186
Db 41 LysAspLeuAlaLysThrAsnGluLeuAspPheTyrProGlyAlaThrHisVal 60
Qy 187 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAATGTGAAGCC 246
Db 61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80
Qy 247 CATTAAATGTGACGGAATTCATGTCAGTCTCTGTCGAGCAGTGGAGATCTTATT 306
Db 81 AlaLeuAspGlnAsnLysMetHisTyrGluLeuLeuHisAspLeuGlnGluLeu 100
Qy 307 CAACAGCAGATTTCC---AAGCAGACAGTCAGCCCCGAGCCTCCGATCGTACTATGAA 363
Db 101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer-----TyrAla 117
Qy 364 CAGTATCACTCACTAAATGAATCTATCTTGGATAGAATTTATACTGAGAGCATCT 423
Db 118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137
Qy 424 GATATGCTTCAAAAAATCCACATTTGATCTCTATTTCGAGAGTACCCACTCTATGTTTA 483
Db 138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157
Qy 484 AAGTTTTCTGAAAGAACAAACAGCCAAAAATGCCATATGGAATGATCTGGAATCCAT 543
Db 158 LysIle---GlyGluLysAsnGluArgLysAlaIlePheMetAspCysGlyIleHis 176
Qy 544 GCCAGAAATGGATCTCTCTGCTTTCTGCTTGGTTCTATA-----585

```

```

Db 177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196
Qy 585 -----585
Db 197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216
Qy 586 -----CGCCAT-----AATCGAATGTGAGAAAGAAC 612
Db 217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236
Qy 613 CGTTCTTTCTATCGGAACCAATCATTCATCGGAACAGACCTGAATAGCACTTTGTCTCC 672
Db 237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255
Qy 673 AAACACTGGTGTGAGGAAGTGCATCCAGTCTCATGCTCGGAAACCTACTGTGGACTT 732
Db 256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275
Qy 733 TATCCTGAGTCAGAACCAAGTGAAGGAGGCTAGTCTGTTCTTGTGAGAAATATCAAC 792
Db 276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295
Qy 793 CAGATTAAAGCATACATCAGCATGCTCATCTACTCCAGCATATAGTGTTCATATTC 852
Db 296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315
Qy 853 TATACAGCAAGTAAAGCAAGACCATGAGGAACTGTCTCTAGTACGAGTGAAGCAGTT 912
Db 316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335
Qy 913 CGTGTCTATTGCAAAACTAGTAAATAACACAGTATACATCGGCATCGGCATCGCAGAAC 972
Db 336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354
Qy 973 TTATACCTAGTCTCTGAGTGGGAGGATGATGATCTATGATTTGGGCATCAATATTCG 1032
Db 355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374
Qy 1033 TTTAC-----ATC 1040
Db 375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394
Qy 1041 AAACCACTGTAGAGAGCTTTTGGCGGTGC 1073
Db 395 LysProThrCysArgGluThrMetLeuAlaVal 405

```

RESULT 14

```

; US-08-277-540-6
; Sequence 6, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Dravna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993

```



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-277-540-6

Alignment Scores:
Pred. No.: 1,14e-65 Length: 417
Score: 635.00 Matches: 139
Percent Similarity: 52.31% Conservative: 76
Best Local Similarity: 33.82% Mismatches: 132
Query Match: 32.87% Indels: 65
DB: 1 Gaps: 9

US-09-980-881a-1_copy_18_1097 (1-1080) x US-08-277-540-6 (1-417)

QY 16 CTTGAGTCCCTTGACCAATGTTCTCTCTGTCGAGCAGCATGCTTCGCG- 66
Db 1 MetArgLeuLeuProValGlyLeuLeuAlaThrThrLeuAlaIleAlaProValArg 20

QY 67 TTCCAGAGTGCCCAAGTTTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db 21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

QY 127 CAGAACTTACTACACATAGATGTTCTCTGCGACGCGGTAAACAGCTGACCTATT 186
Db 41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTrpGlyAlaThrHisHisVal 60

QY 187 GTGAAGAAAAACAAGTCCATTTTTGTAAATGCATCTGATGTCGACAAATGTGAAGCC 246
Db 61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

QY 247 CATTAAATGTGACGGAATTCATGTCAGTGTCTTGTGCGACAGCGTGAAGATCTTATT 306
Db 81 AlaLeuAspGlnAsnLysMetHisTyxGluLeuLeuHisAspLeuGlnGluLeu 100

QY 307 CAACAGCGATTTCC---AAGCACAGTACGCCCCGAGCTCCGCGATCGTACTATGAA 363
Db 101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer-----TyrAla 117

QY 364 CAGTATCACTCACTAAATGAATTTCTTGTGATAGATTTATACTGAGAGCATCCT 423
Db 118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTrp 137

QY 424 GATATGCTTCAAAAATCCATTTGATTCCTTCATTGTGAGAAGTACCCCACTCTATGTTTAA 483
Db 138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyxValLeu 157

QY 484 AAGTTTCTGAAAGAACAAACAGCAAAATGCCATATGATGATGATGATGATGATGAT 543
Db 158 LysIle---GlyGluLysAsnGluArgLysAlaIlePheMetAspCysGlyIleHis 176

QY 544 GCCAGAGAATGGATCTCTCTGCTTCTGCTTGTGTTGTTGTTGTTGTTGTTGTTGTT 585
Db 177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyxGlnAlaThrLysThr 196

QY 585 ----- 585

; Sequence 6, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944

197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyxIleLeuPro 216
586 -----CGCCAT-----AATCGAATGTGAGAAAGAAC 612
217 ValPheAsnValAspGlyTyxIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236
613 CGTTCTTTCTATCGGAACAAATCATTCGATCGGACAGACCTGAATAGCACTTTGTCTCC 672
237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255
673 AAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGACTT 732
256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyxArgGlySer 275
733 TATCTCTGAGTCAGAACCAAGTGAAGGAGTGGCTAGTGTCTTGTGAGAAGAAATATCAAC 792
276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295
793 CAGATTAAAGCATACATCAGCATGCATCTATCTCCAGCATATAGTGTTCATATATTC 852
296 GluIleLysValTyxIleThrPheHisSerTyxSerGlnMetLeuLeuPheProTyxGly 315
853 TATACAGAAAGTAAACAAAGACCATGAGAACTGTCTCTAGTACGACCTGAAGCAATT 912
316 TyxThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335
913 CGTGTCTATTGACAAACTAGTAAAAATACCAGGTATACACATGCCATGCCCTCAGAAACC 972
336 AspValLeu---SerThrArgTyxGluThrArgTyxIleTyxGlyProIleGluSerThr 354
973 TTATACATAGTCTCTGAGGTGGGACGATGATCTATGATTTGGGCATCAATATTCG 1032
355 IleTyxProIleSerGlySerSerLeuAspTrpAlaTyxAspLeuGlyIleLysHisThr 374
1033 TTTAC-----ATC 1040
375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394
1041 AAACCCACCTGTAGAGAGCTTTTGGCGTGC 1073
395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 15
US-08-430-787A-6
; Sequence 6, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
```

```
; FILING DATE: 14-OCT-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Haak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-430-787A-6

Alignment Scores:
Pred. No.: 1.14e-65 Length: 417
Score: 635.00 Matches: 139
Percent Similarity: 52.31% Conservative: 76
Best Local Similarity: 33.82% Mismatches: 132
Query Match: 32.87% Indels: 65
DB: 1 Gaps: 9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-430-787A-6 (1-417)
QY 16 CTTCGACCTCTGTACCCATTGTTCTCTCTGTGAGCAGCATGCTCTCGCG----- 66
Db 1 MetArgLeuLeuValGlyLeuLeuAlaThrThrLeuAlaLeuAlaProValArg 20
QY 67 TTCACAGAGTGGCCAAAGTTCTAGTCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db 21 PheAspArgGluValPheArgValLysProGlnAspGluLysGlnAlaAspIle 40
QY 127 CAGAACTTCTACCAACATAGAAATGTTCTCTGCGCGCGGTAAACAGCTGACCTTATT 186
Db 41 LysAspLeuAlaLysThrAsnGluLeuAspPheTyrProGlyAlaThrHisVal 60
QY 187 GTGAGAGAAAACAAAGTCCATTTTGTAAATGCATCTCATGTCGCAATGTGAAGCC 246
Db 61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGln 80
QY 247 CATTAAATGTGACGGAATTCATGTCAGTCTCTGTCGACAGCGTGAAGATCTTATT 306
Db 81 AlaLeuAspGlnAsnLysMetHisTyrGluLeuLeuLeuHisAspLeuGlnGluLeu 100
QY 307 CAACAGCAGATTTCC---AAACGACACAGTCAGCCCCCGAGCCTCCGCACTGTAATGAA 363
Db 101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer-----TyrAla 117
QY 364 CAGTATCACTCACTAAATGAATCTATCTTGGATAGAATTTATACTGAGAGCATCTT 423
Db 118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137
QY 424 GATATGCTTCAAAAATCCACATGGATCTCTATTTCGAGAGTACCCACTCTATGTTTA 483
Db 138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157
QY 484 AGGTTTCTGAAAAGAACAAACAGCCAAAATATGCCATATGGATTGATGGAATCCAT 543
Db 158 LysIle---GlyGluLysAsnGluArgLysAlaIlePheMetAspCysGlyIleHis 176
QY 544 GCCAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 585
Db 177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196
QY 585 ----- 585
Db 197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216
```

```
QY 586 -----CGCCAT-----AATCGAATGTGAGAAAGAAC 612
Db 217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236
QY 613 CGTTCTTTCTATCGGAACAAATCATTCATCGGAACAGACCTGAATAGCACTTTGTCTCC 672
Db 237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255
QY 673 AAACACTGTGTGAGGAAGTGCATCCAGTCTCTCATGCTCGGAACCTACTCTGGAATT 732
Db 256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275
QY 733 TATCCTGAGTCAGAACCAAGTGAAGCAGCTGGTAGTCTTCTTGAGAAGAAATATCAAC 792
Db 276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295
QY 793 CAGATTAAAGCATACATCAGCATGCTCATCTACTCCAGCATATAGTGTTCATATTC 852
Db 296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315
QY 853 TATACACGAAGTAAAGCAAGACCATGAGAACTCTCTCTAGTACCCAGTGAAGCAGTT 912
Db 316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335
QY 913 CGTGTATTGCAAAACTAGTAAATACCAAGTATACATCGCCATGCGCTCAGAAACC 972
Db 336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354
QY 973 TTATACCTAGTCTCTCGAGTGGGAGCATTTGGATCTATGATTTGGGCATCAATATTCG 1032
Db 355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374
QY 1033 TTTTAC-----ATC 1040
Db 375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394
QY 1041 AAACCCACCTGTAGAGAAGCTTTTGGCGCTGTC 1073
Db 395 LysProThrCysArgGluThrMetLeuAlaVal 405
```

Search completed: January 1, 2004, 19:03:29

Job time : 50.1067 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:59:01 ; Search time 47.8326 Seconds
(without alignments)
9060.445 Million cell updates/sec

Title: US-09-980-881a-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagcttgagccttgc.....ttttgcgcgtctctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 1467874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2 1/USPTO spoof/US09980881/runat 31122003 135753 25135/app query.fasta_1.2958
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09980881 @CGN 1 1 148 @runat 31122003 135753 25135
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2 6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2 6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2 6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2 6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2 6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2 6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2 6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2 6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2 6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2 6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2 6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2 6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2 6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2 6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2 6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2 6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2 6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2 6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1835	95.0	423	12	US-10-379-836-17	Sequence 17, Appl
2	1746	90.4	423	12	US-10-379-836-2	Sequence 2, Appl1
3	1550.5	80.3	422	12	US-10-379-836-18	Sequence 18, Appl
4	1522.5	78.8	422	12	US-10-379-836-16	Sequence 16, Appl
5	839	43.4	211	9	US-09-925-302-467	Sequence 467, App
6	635	32.9	417	12	US-10-341-434-188	Sequence 188, App
7	635	32.9	417	15	US-10-229-546-2	Sequence 2, Appl1
8	635	32.9	417	15	US-10-229-546-9	Sequence 9, Appl1
9	632	32.7	402	12	US-10-379-836-20	Sequence 20, Appl
10	623	32.2	716	10	US-09-910-059-125	Sequence 125, App
11	613.5	31.8	437	12	US-10-274-639-12	Sequence 12, Appl
12	613.5	31.8	437	14	US-10-200-344-10	Sequence 10, Appl
13	564.5	29.2	374	9	US-09-888-615-61	Sequence 61, Appl
14	563	29.1	444	16	US-10-176-306-74	Sequence 74, Appl
15	560	29.0	613	10	US-09-910-059-113	Sequence 113, App
16	546	28.3	399	15	US-10-200-910-8	Sequence 8, Appl1
17	528.5	27.4	350	14	US-10-200-344-12	Sequence 12, Appl
18	519.5	26.9	436	15	US-10-200-910-6	Sequence 6, Appl1
19	511.5	26.5	436	14	US-10-076-535-2	Sequence 2, Appl1
20	508	26.3	417	9	US-09-923-779-150	Sequence 150, App
21	504	26.1	428	9	US-09-925-297-528	Sequence 528, App
22	495.5	25.6	298	12	US-10-379-836-19	Sequence 19, Appl
23	489.5	25.3	406	9	US-09-925-297-517	Sequence 517, App
24	489.5	25.3	419	12	US-10-345-680-65	Sequence 65, Appl
25	483	25.0	421	11	US-09-946-374-234	Sequence 234, App
26	483	25.0	421	12	US-10-015-387A-234	Sequence 234, App
27	483	25.0	421	12	US-10-006-130A-234	Sequence 234, App
28	483	25.0	421	12	US-10-199-672-308	Sequence 308, App
29	483	25.0	421	12	US-10-006-172A-234	Sequence 234, App
30	483	25.0	421	12	US-10-187-749-308	Sequence 308, App
31	483	25.0	421	12	US-10-194-457-308	Sequence 308, App
32	483	25.0	421	12	US-10-184-642-308	Sequence 308, App
33	483	25.0	421	12	US-10-196-747-308	Sequence 308, App
34	483	25.0	421	12	US-10-015-392A-234	Sequence 234, App
35	483	25.0	421	12	US-10-017-253A-234	Sequence 234, App
36	483	25.0	421	12	US-10-173-690-308	Sequence 308, App
37	483	25.0	421	12	US-10-173-692-308	Sequence 308, App
38	483	25.0	421	12	US-10-173-693-308	Sequence 308, App
39	483	25.0	421	12	US-10-173-694-308	Sequence 308, App
40	483	25.0	421	12	US-10-173-698-308	Sequence 308, App
41	483	25.0	421	12	US-10-173-699-308	Sequence 308, App
42	483	25.0	421	12	US-10-173-707-308	Sequence 308, App
43	483	25.0	421	12	US-10-174-569-308	Sequence 308, App
44	483	25.0	421	12	US-10-174-583-308	Sequence 308, App
45	483	25.0	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Alignment Scores:	1.19e-176	Length:	423
Pred. No.:	1835.00	Matches:	356
Score:			

Percent Similarity: 86.23% Conservative: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 12 Gaps: 2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-17 (1-423)

QY 1 ATGAAGCTTTGCGAGCTTGTGAGTCCCTGTTGACCCATGCTCTCTCTGAGCAGCATGTC 60
DB 1 MetLysLeuCysserLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 61 TTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 120
DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 121 GTTCTACAGAACTTACTACAAATGAGATTGTTCTCTGGCGCCGCGTAACAGCTGAC 180
DB 41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 181 CTTATTCTGAAGAAAACAAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 240
DB 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCACTGCTTGTCTGGCAGACGTCGGAAGAT 300
DB 81 LysAlaHisLeuAsnValSerGlyIleProCysserValLeuLeuAlaAspValGluAsp 100
QY 301 CTTATTCAACAGCAGATTTCCACGACACACAGTACGCCCCGAGCTCCGATGCTACTAT 360
DB 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
QY 361 GACAGTATCACTACATAAGTAATCTATCTTGTGATAGATTTATTAACGTGAGGCGAT 420
DB 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
QY 421 CCTGATATGCTTCAAAAAATCCCATTTGAGTCTCTCTTTCAGAGTACCCCACTCTATGTT 480
DB 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
QY 481 TTAAGCTTTCTGAAAAGAACAAACAGCCAAAATGCCATATGATGATGATGCGAATC 540
DB 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
QY 541 CATGCCAGAGATGATCTCTCTGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 591
DB 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
QY 591 ----- 591
DB 201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
QY 592 -----AATCGAATGTGAGAAAG 609
DB 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 610 AACGGTTCTTCTATGCGAACATCATTGATCGGACAGACCTGATAGCACTTTGTC 669
DB 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 670 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTATGCTCGGAAACCTACTGTGGA 729
DB 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280
QY 730 CTTTATCCTGAGTACAGAACAGAGAGTGAAGCAGTGTGCTAGTTCTTTGAGAAGAAATATC 789
DB 281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
QY 790 AACGAGATTAAGCATACATCAGATGATTCATCTACTCCAGCATATAGTTTCCATAT 849
DB 301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 850 TCCTATACAGAGTAAAGCAAGACCATGAGAACTGCTCTAGTAGCAGTGAAGCA 909

DB 321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
QY 910 GTTCTGCTATTACAAAATAGTAAATACCAAGTATACATGCGCATGCTCAGAA 969
DB 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360
QY 970 ACCTTATACCTAGCTCTCTGAGGTGGGACGATTGGATCTATGATTGGGCATCAAAATAT 1029
DB 361 ThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1030 TCCTT-----TAC 1037
DB 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400
QY 1038 ATCAAAACCCACCTGTAGAGAGCTTTTCCGCTGCTCTCTAA 1079
DB 401 IleLysProThrCysArgGluAlaPheAlaValSerLys 414

RESULT 2
US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Papio hamadryas
US-10-379-836-2

Alignment Scores:
Pred. No.: 1,2e-167 Length: 423
Score: 1746.00 Matches: 338
Percent Similarity: 83.82% Conservative: 9
Best Local Similarity: 81.64% Mismatches: 13
Query Match: 90.37% Indels: 55
DB: 12 Gaps: 2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-2 (1-423)

QY 1 ATGAAGCTTTGCGAGCTTGTGAGTCCCTGTTGACCCATGCTCTCTCTGAGCAGCATGTC 60
DB 1 MetLysLeuCysserLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
QY 61 TTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 120
DB 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
QY 121 GTTCTACAGAACTTACTACAAATGAGATTGTTCTCTGGCGCCGCGTAACAGCTGAC 180
DB 41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
QY 181 CTTATTCTGAAGAAAACAAAGTCCATTTTGTAAATGCATCTGATGTCGCAATGTG 240
DB 61 LeuIleValLysLysGlnValHisPhePheValAsnSerSerAspValAspAsnVal 80
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCACTGCTTGTCTGGCAGACGTCGGAAGAT 300
DB 81 LysAlaHisLeuAsnValSerGlyIleProCysserValLeuLeuAlaAspValGluAsp 100
QY 301 CTTATTCAACAGCAGATTTCCACGACACACAGTACGCCCCGAGCTCCGATGCTACTAT 360
DB 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

```
QY 361 GAACAGTATCACTCAATAATGAATCTATCTTGGATAGATTATTAACAGAGCAT 420
Db 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluLeuIleThrGluLysTyr 140
QY 421 CCTGATATGCTTACAAAATCCCAATGGATCTCATTTGAGAAAGTACCCACTCTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerTyrGluLysHisProLeuTyrVal 160
QY 481 TTTAAGGTTTCTGGAAGAACAACACCAACCAAAATGCCATATGATGATGCTGTGGAATC 540
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaMetTrpIleAspCysGlyIle 180
QY 541 CATTGCCAGAGATCGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 591
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGlu 200
QY 591 ----- 591
Db 201 TyrTyrGlyIleIleGlyGluTyrThrAsnLeuLeuArgHisValAspPheTyrValMet 220
QY 592 -----AATCGAATGTGGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
QY 610 AACCGTTCTTCTATGCGAACAATCATTCATCGGAACACACCTGAATAGCAACTTTGTC 669
Db 241 AsnArgSerPheTyrAlaAsnAsnArgCysIleGlyThrAspLeuAsnArgAsnPheAla 260
QY 670 TCCAAACACTGGTGTAGGAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 280
QY 730 CTTTATCCTGAGTCAGAACACAGAGTGAAGCGAGTGTCTTCTTGAAGAATAATC 789
Db 281 LeuTyrProGluSerGluProGluAlaLysAlaValAlaAsnPheLeuArgAsnIle 300
QY 790 AACGAGTTAAGCATACATCAGATCATTCATCTATCTCCAGACATATAGTGTTCATAT 849
Db 301 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320
QY 850 TCCTATACAGCACTAAAGCAAGACCATGAGCACTGTCTAGTAGCAGTGAAGCA 909
Db 321 SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 340
QY 910 GTTCGTCTATTGCAAAAACCTAGTAAATAACCAAGTATACATGCGCCATGCTCAGAA 969
Db 341 ValArgAlaIleGlnLysThrSerLysAsnIleArgTyrThrHisGlyArgGlySerGlu 360
QY 970 ACCTTATACCTAGCTCTGAGGAGTGGGAGCATTTGATCTATGATTTGGGCATCAATAT 1029
Db 361 ThrLeuTyrLeuAlaProGlyGlyAlaAspAspTrpIleTyrAspLeuGlyIleLysTyr 380
QY 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyLysTyrGlyPheLeuLeuProGluArgTyr 400
QY 1038 ATCAAAACCCACCTGTAGAGAGCTTTTGGCCGCTCTCTCTAAA 1079
Db 401 IleLysProThrCysLysAspAlaPheAlaValSerLys 414
```

RESULT 3

```
US-10-379-836-18
; Sequence 18, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
```

```
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-379-836-18

Alignment Scores:
Pred. No.: 7,23e-148 Length: 422
Score: 1550.50 Matches: 302
Percent Similarity: 78.99% Conservatives: 25
Best Local Similarity: 72.95% Mismatches: 32
Query Match: 80.25% Indels: 56
DB: 12 Gaps: 3
```

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-18 (1-422)

```
QY 1 ATGAAGCTTTGGAGCCTTGCAGCTTGTACCCATTGTTCTTCTGTGAGCAGCATGTC 60
Db 1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr---GluGlnHisGly 19
QY 61 TTCGCGTTCCAGAGTGGCCAAAGTTCTTAGCTGTCTTCTAGAACCTCTAGCAAGTTCAA 120
Db 20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39
QY 121 GTTCTACAGAAATCTTACTACACATATGAGATTGTTCTCTGGCAGCGCGTAACAGCTGAC 180
Db 40 LeuLeuGlnAsnLeuThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59
QY 181 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAATGTG 240
Db 60 PheIleGluLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79
QY 241 AAGAGCCATTAAATGTGCGCAATTCATGTCAGTCTCTGCGCAGCGCTCGCATCTAT 300
Db 80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnValGluAsp 99
QY 301 CTTATTCAACAGCAGATTTTCCAAACAGACAGTCCAGCCCCGAGCCTCGCATCTAT 360
Db 100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119
QY 361 GAACAGTATCACTCAATAATGAATCTATCTTGGATAGATTATTAACAGAGCAT 420
Db 120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139
QY 421 CCTGATATGCTTACAAAATCCCAATGGATCTCATTTGAGAAAGTACCCACTCTATGTT 480
Db 140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159
QY 481 TTTAAGGTTTCTGGAAGAACAACACCAACCAAAATGCCATATGATGATGCTGTGGAATC 540
Db 160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179
QY 541 CATTGCCAGAGATCGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 591
Db 180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199
QY 591 ----- 591
Db 200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219
QY 592 -----AATCGAATGTGGAGAAAG 609
Db 220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239
QY 610 AACCGTTCTTCTATGCGAACAATCATTCATCGGAACACACCTGAATAGCAACTTTGTC 669
Db 240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259
QY 670 TCCAAACACTGGTGTAGGAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db 260 SerLysHisTrpCysGluLysGlyAlaSerSerCysSerGluThrTyrCysGly 279
```



```
Qy 585 ----- 585
Db 197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216
Qy 586 -----GGCCAT-----AATCAATGTGGAGAAAGAAC 612
Db 217 ValPheAsnValAspGlyTyrIleTyrSerTrpThrLysAsnArgMetTrpArgLysAsn 236
Qy 613 CGTTCTTCTATGCGAACAATCATTCATCGGACACACCTGATACCAACTTGTCTCC 672
Db 237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255
Qy 673 AAACACTGGTGTGAGGAAGGTGATCCAGTCTCTCATGCTCGGAAACCTACTGTGGACTT 732
Db 256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275
Qy 733 TATCTGAGTCAGAACAGAGTGAAGCGAGTGCTAGTTCTTGTGAGAGAAATATCAAC 792
Db 276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295
Qy 793 CAGATTAAAGCATACATCAGCATTCATCTACTCCAGCATATAGTGTTCATATCC 852
Db 296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315
Qy 853 TATACAGCAAGTAAAGCAAGACCATCAGGAATCTGTCTAGTAGCCAGTGAAGCAGTT 912
Db 316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335
Qy 913 CGTGCTATTGACAAAACCTAGTAAATAACAGGATATACATGCGCCATGCTCAGAAACC 972
Db 336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354
Qy 973 TTATACCTAGTCTCTGAGGTGGGAGCATGATGATCTATGATTTGGGCATCAATATCG 1032
Db 355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374
Qy 1033 TTATAC-----ATC 1040
Db 375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394
Qy 1041 AAACCCACCTGTAGAGAGCTTTTGGCGGTGC 1073
Db 395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 7
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MP101-156P1RN
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-546-2

Alignment Scores:
Pred. No.: 3,04e-55 Length: 417
Score: 635.00 Matches: 139
Percent Similarity: 52.31% Conservative: 76
Best Local Similarity: 33.82% Mismatches: 132
Query Match: 32.87% Indels: 65
```

```
DB: 15 Gaps: 9
US-09-980-881a-1_copy_18_1097 (1-1080) x US-10-229-546-2 (1-417)
Qy 16 CTTGCACTCTGTACCACTTGTCTCTGTGAGCAGCATGCTTCGCG----- 66
Db 1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20
Qy 67 TTCAGAGTGGCCCAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAGTCTTA 126
Db 21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40
Qy 127 CAGAACTTCTACTCAACATATGAGATTGTCTCTGCGAGCCGCGTAACAGCTGACCTTAT 186
Db 41 LysAspLeuAlaLysThrAsnGluLeuAspPheTyrTrpProGlyAlaThrHisVal 60
Qy 187 GTGAAGAAAAACAAGTCCATTTTGTAAATGTCATCTGATGTGCACAAATGTGAAGCC 246
Db 61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80
Qy 247 CATTAAATGTGAGCGAATTCCATGCACTGCTCTGTCGACAGCTGGAAGATCTTATT 306
Db 81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluIle 100
Qy 307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCGAGCCCTCCGCATCGTACTATGAA 363
Db 101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer-----TyrAla 117
Qy 364 CAGTATCACTACTAAATGAAATCTATTCTTGGATAGAATTTTATACTGAGAGCATCT 423
Db 118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137
Qy 424 GATATCTTACAAAATCCATGTCATCTCTATTGAGAAGTACCACCTCTATGTTTAA 483
Db 138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157
Qy 484 AAGGTTTCTGGAAGAAACAAACAGCCAAATGTCATATGATGTCATGTCGAATCCAT 543
Db 158 LysIle---GlyLysLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176
Qy 544 GCAGAGAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 585
Db 177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196
Qy 585 ----- 585
Db 197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216
Qy 586 -----GGCCAT-----AATCAATGTGGAGAAAGAAC 612
Db 217 ValPheAsnValAspGlyTyrIleTyrSerTrpThrLysAsnArgMetTrpArgLysAsn 236
Qy 613 CGTTCTTCTATGCGAACAATCATTCATCGGACACACCTGATACCAACTTGTCTCC 672
Db 237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255
Qy 673 AAACACTGGTGTGAGGAAGGTGATCCAGTCTCTCATGCTCGGAAACCTACTGTGGACTT 732
Db 256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275
Qy 733 TATCTGAGTCAGAACCAAGTGAAGCGAGTGGCTAGTTCTTGTGAGAGAAATATCAAC 792
Db 276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295
Qy 793 CAGATTAAAGCATACATCAGCATTCATCTACTCCAGCATATAGTGTTCATATCC 852
Db 296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315
Qy 853 TATACAGCAAGTAAAGCAAGACCATGAGGAATCTGTCTAGTAGCCAGTGAAGCAGTT 912
Db 316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335
Qy 913 CGTGCTATTGACAAAACCTAGTAAATAACAGGATATACATGCGCCATGCTCAGAAACC 972
```

```
Db 336 AspValLeu---SerThrArgTy-GluThrArgTyrIleTyArgProIleGluSerThr 354
Qy 973 TTATACCTAGCTCTCGAGGAGGCGGACGATTGGATCTATGATTTGGGCATCAAAATATTCG 1032
Db 355 IleTyProIleSerGlySerSerLeuAspTrpAlaTyAspLeuGlyIleLysHisThr 374
Qy 1033 TTTAC-----ATC 1040
Db 375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394
Qy 1041 AAACCCACCTCTAGAGAGCTTTTGGCGCTGC 1073
Db 395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 8
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Welch, Nadine S.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Carroll, Joseph M.
; TITLE OF INVENTION: 6299. A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MP101-156P1RNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; PRIOR FILING DATE: 2002-08-28
; PRIOR FILING DATE: 60/316,575
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-546-9

Alignment Scores:
Pred. No.: 3,04e-55 Length: 417
Score: 635.00 Matches: 139
Percent Similarity: 52.31% Conservative: 76
Best Local Similarity: 33.82% Mismatches: 132
Query Match: 32.87% Indels: 65
DB: 15 Gaps: 9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-229-546-9 (1-417)
Qy 16 CTTCAGTCTTGTTGACCATGTTCTCTCTGTGAGCAGCATGCTTCGCG----- 66
Db 1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20
Qy 67 TTCCAGAGTGCCCAAGTCTAGCTCTCTCTAGAACCTCTAGGCAGTTCAGTTCTA 126
Db 21 PheAspArgGluTyValPheArgValLysProGlnAspGluLysGlnAlaAspIle 40
Qy 127 CAGAACTCTTACTCAACATATGAGATGTTCTCTGCGAGCGCGTAACAGCTGACCTATT 186
Db 41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyProGlyAlaThrHisHisVal 60
Qy 187 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGTATGTCGACAAATGTGAAGCC 246
Db 61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80
Qy 247 CATTTAATGTGCGGAAATTCATGCGTCTCTGCGCAGCAGCTGGAAGATCTATT 306
Db 81 AlaLeuAspGlnAsnLysMetHisTyArgIleLeuIleHisAspLeuGlnGluIle 100
Qy 307 CAACAGCAGATTTC---AAGCAGACAGTCAGCCCGAGCTCCGCGATCTGCTACTATGAA 363
Db 101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer-----TyrAla 117
Qy 364 CAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAATTTATAACTGAGAGCATCCT 423
```

```
Db 118 LysTyAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyPro 137
Qy 424 GATATGCTTACAAAAATCCACATTTGGATCTCTCTTTCGAGAAGTACCCACTCTATGTTTAA 483
Db 138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyValLeu 157
Qy 484 AAGTTTCTCGAAGAAACAAACAGCCAAATATGCCATATGGATTGATGCTGGGAATCCAT 543
Db 158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176
Qy 544 GCCAGAGAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 585
Db 177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyGlnAlaThrLysThr 196
Qy 585 ----- 585
Db 197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyIleLeuPro 216
Qy 586 -----GCCCAT-----AATCGAATGTGGAGAAAGAAC 612
Db 217 ValPheAsnValAspGlyTyIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236
Qy 613 CGTCTCTTCTATCGGAACAATCATTCGATCGGAACACACCTGATAGCACTTTGTCTCC 672
Db 237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255
Qy 673 AAACACTGTGTGAGGAAGTGTCATCCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 732
Db 256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyArgGlySer 275
Qy 733 TATCTCTGAGTCAGAACCAAGTGAAGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 792
Db 276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295
Qy 793 CAGATTAAACATACATCAGCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 852
Db 296 GluIleLysValTyIleThrPheHisSerTySerGlnMetLeuLeuPheProTyArg 315
Qy 853 TATACGCAAGTAAAGCAAGACCATGAGCACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 912
Db 316 TyrThrSerLysLeuProProAsnHisGluLysLeuAlaLysValAlaLysIleGlyThr 335
Qy 913 CGTGTATTGACAAACTAGTAAAAATACCAGGTATACATACATGCGCATGCTCAGAAACC 972
Db 336 AspValLeu---SerThrArgTyArgGluThrArgTyIleTyArgProIleGluSerThr 354
Qy 973 TTATACCTAGCTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATATTCG 1032
Db 355 IleTyProIleSerGlySerSerLeuAspTrpAlaTyAspLeuGlyIleLysHisThr 374
Qy 1033 TTTAC-----ATC 1040
Db 375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394
Qy 1041 AAACCCACCTGTAGAGAGCTTTTGGCGCTGC 1073
Db 395 LysProThrCysArgGluThrMetLeuAlaVal 405
```

RESULT 9

```
US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR FILING DATE: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-20

Alignment Scores:

Pred. No.: 6, 06e-55 Length: 402
Score: 632.00 Matches: 134
Percent Similarity: 52.31% Conservative: 70
Best Local Similarity: 34.36% Mismatches: 126
Query Match: 32.71% Indels: 61
DB: 12 Gaps: 7

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-20 (1-402)

```
Qy 67 TTCAGAGTGGCCAAAGTTCTAGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTTA 126
Db 7 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisLeile 26
Qy 127 CAGAACTTACTACAAATGATGATGCTTCTGCGAGCGGTAACAGCTGACCTTATT 186
Db 27 ArgGluLeuAlaSerThrThrGlnleAspPheTrpLysProAspSerValThrGlnle 46
Qy 187 GTGAAGAAACAAAGTCCATTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
Db 47 LysProHisSerThrValAspPheArgValLysValGluAspThrValThrValGluAsn 66
Qy 247 CATTTAAATGTGAGCGGAATTCATGCGAGTCTTGTGCGAGAGTGGAGATCTTATT 306
Db 67 ValLeuLysGlnAsnGluLeuGlnTyrlsValLeuileSerAsnLeuArgAsnValVal 86
Qy 307 CAACAGCAGATTTCCACGACACAGTACGCCCCGAGCTCCGATCGCTATGAAACAG 366
Db 87 GluAlaGlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrlsGluLys 103
Qy 367 TATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTAGAGCATCTTGAT 426
Db 104 TyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAla 123
Qy 427 ATGCTTCAAAATCCACATGGATCTCTCAATTCAGAAAGTACCCACATCTATGTTTTAAAG 486
Db 124 LeuileSerArgSerValIleGlyThrPheGluGlyArgAlaIleTyrlsLeuLys 143
Qy 487 GTTCTGGAAGAACCAACAGCCAAAGATGATGATGATGATGATGATGATGATGATGCC 546
Db 144 Val---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAla 162
Qy 547 AGAAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 585
Db 163 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyr 182
Qy 585 ----- 585
Db 183 GlyArgGluIleGlnValThrGluLeuAsnLysLeuAspPheTyrlsValLeuProVal 202
Qy 586 -----GCCAT-----AATCAATGTGGAAGAACCGT 615
Db 203 LeuAsnleAspGlyTyrlsThrTrpThrLysSerArgPheTrpA:glYsThrArg 222
Qy 616 TCTTCTATCGAACAATCATGATCGGACAGACCTGATAGCACTTGTCTCCAAA 675
Db 223 SerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAla 241
Qy 676 CACTGGTGTGAGGAAGTGCATCCAGTCTCTGTCGGAACCTACTGTGGACTTTAT 735
Db 242 GlyTrpCysGluIleGlyAlaSerArgAsnProCysaspGluThrTyrlsCysGlyProAla 261
Qy 736 CTTGAGTCAGAACGAGTGAAGGCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 795
Db 262 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSer 281
```

```
Qy 796 ATTAAGACATACATCAGCATGCTATCATATCCAGCATATAGTGTTCATATTCCTAT 855
Db 282 IleLysAlaTyrlsLeuThrIleHisSerTyrlsSerGlnMetMetIleTyrlsProTyrlsSerTyr 301
Qy 856 ACACGAAGTAAAGCAAGACCACTAGAGAACTGCTCTCTAGTAGCAGTGAAGAGTTTCGT 915
Db 302 AlaTyrlsLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLysAlaThrValLys 321
Qy 916 GCTATTGACAAAATACTAGTAAAAATACCAAGTATACATGCGCATGCGCTCAGAAACCTTA 975
Db 322 GluLeu---AlaSerLeuHisGlyThrTyrlsThrTyrlsGlyProGlyAlaThrThrile 340
Qy 976 TACCTAGCTCTGAGAGTGGGAGCATGATGATCTATGATTTGGCATCAATATTCGTTT 1035
Db 341 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrlsPheGlnGlyIleArgTyrlsSerPhe 360
Qy 1036 ACATC-----AAA 1043
Db 361 ThrPheGluLeuArgAspThrGlyArgTyrlsGlyPheLeuLeuProGluSerGlnIleArg 380
Qy 1044 CCACCTGTAGAGAGCTTTTGGCGGTGTC 1073
Db 381 AlaThrCysGluGluThrPheLeuAlaIle 390
RESULT 10
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; FILE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-125
```

Alignment Scores:
Pred. No.: 5, 8e-54 Length: 716
Score: 623.00 Matches: 137
Percent Similarity: 51.34% Conservative: 74
Best Local Similarity: 33.33% Mismatches: 139
Query Match: 32.25% Indels: 62
DB: 10 Gaps: 8

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-910-059-125 (1-716)

```
Qy 7 CTTTGACGCTTGCAGTCTCTGATCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 66
Db 1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisGlyGlyGlu 20
Qy 67 ---TTCAGAGTGGCAAGTTCTAGTGTCTTCTTCTAGAACCTCTAGGCAAGTTCAAGTT 123
Db 21 HisPheGluGlyGluLysValPheArgValAsnValGluAsnHisIleAsnIle 40
```


Db 169 TyrGluGlyArgSerLeuPheLeuLysLeu---GlyArgArgSerArgLeuLysArg 187
Qy 517 GCCATATGAGTACTGTGAATCCATGCCAGAGATGGATCTCTCTGCTTTCTGCTTG 576
Db 188 AlaValTrpIleAepCysGlyIleHisAlaArgGluTrpIleGlyProAlaPheCysGln 207
Qy 577 TGGTTTCATA----- 585
Db 208 TrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArgLysMetLeu 227
Qy 585 ----- 585
Db 228 AsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHisPheSerTrp 247
Qy 586 GGCCATAATCGAATGTCGAGAAAGACCGTTCTTTCATCGCAACAATCATTCATCGGA 645
Db 248 ThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArgCysArgGly 267
Qy 646 ACAGACCTGATACCACTTTGCTCCAAACACTGGTGTGAGNAGGTGCATCCAGTTCC 705
Db 268 ValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAlaSerMetHis 286
Qy 706 TCATGCTCGGAAACCTACTGTGAGCTTTATCTCAGTCAGAACAGAACGAGGAGGAGTG 765
Db 287 ProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluValLysAlaVal 306
Qy 766 GCTAGTTCTTGAAGAAATATCAACAGATTAAGCATACATCAGCATGCAATTCATAC 825
Db 307 AlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPheHisAlaTyr 326
Qy 826 TCCAGCATATAGTGTTCATATCTCATACGAAGTAAAGCAAGACCATGAGGAA 885
Db 327 AlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsnPheArgCys 346
Qy 886 CTGCTCTAGTAGCAGTGAAGCAGTTCGTCCTATTGACAAACTAGTAAATACAGG 945
Db 347 ValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyrGlyValArg 365
Qy 946 TATACATGCGCCATGCTCAGAAACCTTATACATGCTCTGCTGAGTGGGAGCATGG 1005
Db 366 TyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSerMetAspTrp 385
Qy 1006 ATCTATGATTGGGCATCAATATTCGTTTAC----- 1037
Db 386 AlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThrGlyTyrPhe 405
Qy 1038 -----ATCAACCCACCTGTAGAGAGCTTTTGGCGTGC 1073
Db 406 GlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMetLeuAlaVal 425

RESULT 13

US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARVDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 374
; TYPE: PR1
; ORGANISM: Homo sapiens

US-09-888-615-61

Alignment Scores:

Pred. No.: 4,01e-48 Length: 374
Score: 564.50 Matches: 115
Percent Similarity: 53.70% Conservative: 52
Best Local Similarity: 36.98% Mismatches: 87
Query Match: 29.22% Indels: 58
DB: 9 Gaps: 6

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-888-615-61 (1-374)

Qy 307 CACACAGAGATTTCCAAACGACACAGATCAGCCCCGAGCCCTCCGATCGTAC---TATGAA 363
Db 29 ArgGlnGluIleValAspLysSerValSerTrpSerLeuGluThrTyrSerTyrAsn 48
Qy 364 CAGTATCACCTACATAATGAAATCTATTCTTGATAGAAATTTATACTAGAGAGCATCCT 423
Db 49 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 68
Qy 424 GATATGCTTACAAAAATCCACATTTGGATCCTCTTTTGAGAAAGTACCCACTCTATGTTTA 483
Db 69 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProMetTyrTyrLeu 88
Qy 484 AAGTTTTCGAAAGAAAGAAACAAACAGCCAAAATGCCATATGATGATGATGATGATGAT 543
Db 89 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 108
Qy 544 GCCAGAGATGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 585
Db 109 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 128
Qy 585 ----- 585
Db 129 HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 148
Qy 586 -----GGCCAT-----AATCGAATGTGAGAGAAAGAAC 612
Db 149 ValLeuAsnIleAspGlyTyrIleTyrThrTyrThrTyrThrTyrThrTyrThrTyrThr 168
Qy 613 CGTTCTTCTATCGCAACAATCATTCATCGGACGACCTGATGATGATGATGATGATGATGAT 672
Db 169 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 187
Qy 673 AAACACTGTTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
Db 188 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 207
Qy 733 TATCCTGAGTCAGAACACAGAGTGAAGCAGTGGCTAGTGTCTTGTGAGAGAAATATCAAC 792
Db 208 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 227
Qy 793 CAGATTAAAGCATACATCAGCATGCTTCTATCTATCTCCAGCATATATGTTTCCATATCC 852
Db 228 AspIleLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 247
Qy 853 TATACAGAGTAAAGCAAGACCATGAGAACTGTCTCTAGTACGCTGAGGAGAGT 912
Db 248 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 267
Qy 913 CGTGCTATTGACAAAACTAGTAAAAATACACAGTATACATCGCCATGCTCAGAAACC 972
Db 268 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 286
Qy 973 TTATACCTAGCTCCTCGAGTGGGAGCATGATGATGATGATGATGATGATGATGATGATGAT 1032
Db 287 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspIleGlyIleProPheSer 306
Qy 1033 TTTACATC----- 1040
Db 307 TyrThrPheGluLeuArgAspSerGlyThrTyrGlyPheValLeuProGluAlaGlnIle 326
Qy 1041 AAACCCACCTGTAGAGAGCTTTTGGCGTGC 1073

;; PRIOR APPLICATION NUMBER: GB 9703103.3
;; PRIOR FILING DATE: 1997-02-14
;; PRIOR APPLICATION NUMBER: GB9609405.7
;; PRIOR FILING DATE: 1996-05-04
;; NUMBER OF SEQ ID NOS: 131
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 113
;; LENGTH: 613
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: humanised fd mutant HCPB sequence
US-09-910-059-113

Alignment Scores:
Pred. No.: 1.31e-47 Length: 613
Score: 560.00 Matches: 134
Percent Similarity: 45.11% Conservative: 55
Best Local Similarity: 31.98% Mismatches: 114
Query Match: 28.99% Indels: 117
DB: 10 Gaps: 11

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-910-059-113 (1-613)

```
QY 3 GAAGCTTGGAGCCTTGCAGTCTTGTACCAATGTTCTCTCTGTGAGCAGCATGTCCT 62
Db 238 GluLeuLysThrProLeuGlyAspThrThrHis-----ThrCysPro 251
QY 63 CGCGTTCCAGAGTGGCCAAAGTTCTAGCTGTCTTCTCTAGAACCTCTAGGCAAGTCAAGT 122
Db 252 ArgCysProGlu-----ProLysSerCys----- 259
QY 123 TCTACAGATCTTACTACAACATATGAGATTGTTCTCTGCGAGCGGTAACAGCTGACCT 182
Db 260 -----AspThrProPro 263
QY 183 TATTGTGAAGAAAACAAAGTCCATTTTGT-----AAATGC 221
Db 264 ProCysProArgCysProGluProLysSerCysAspThrProProCysProArgCys 283
QY 222 ATCTGATGTGCACAAATGTGAAGCCCAATTTAAATGTGAGGGAATTCATGCGAGTGTCTT 281
Db 284 ProGluProLysSerCysAspThrProProCysProArg-CysProAlaProGluLe 303
QY 282 GCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCACACGACACAGTCAGCCCCCG 341
Db 303 uLeuGlyGly----- 306
QY 342 AGCTCCGCATCGTACTATGACAGATATCACTACTAAATGAATCTATTCTTGGATAGA 401
Db 307 -AlaThrGlyHisSerTyrGluLysTyrAsnLysTrpGluThrIleGluAlaTrpThrGl 326
QY 402 ATTTATAACTGAGAGGATCTGATATGCTTACAAAATCCACATTCGATCCCTCATTTGA 461
Db 326 nGlnValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGlyThrThrPheGl 346
QY 462 GAAGTACCCACTCTATGTTTAAAGTTTCTGAAAAAGAAACAAACAGCCCAAAATGCCAT 521
Db 346 uGlyArgAlaIleTyrLeuLeuLysVal---GlyLysAlaGlyGlnAsnLysProAlaI 365
QY 522 ATGGAATGATGTGGAAATCCATGCGACAGAAATGATCTCTCTGCTTTCTGCTTGTGCTT 581
Db 365 ePheMetAspCysGlyPheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPh 385
QY 582 CATA----- 585
Db 385 eValArgGluAlaValArgThrTyrGlyArgGluIleGlnValThrGluLeuLeuAspLy 405
QY 586 -----GGCCAT----- 591
Db 405 sLeuAspPheTyrValLeuProValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLy 425
QY 592 -AATCGAATGTGGAGAAAGAACCGTCTTCTTATGCGAACAAATCATTTGCATCGGAACAGA 650
```

```
Db 425 sSerArgPheTrpArgLysThrArgSerThrHisThrGlySerSerCysIleGlyThrAs 445
QY 651 CCTGAATAGCAAACTTTGCTCTCCAAACACACTGGTGTGAGAAAGGTGCTCCAGTTCTCTCATG 710
Db 445 pProAsnArgAsnPhe---AspAlaGlyTrpCysGluIleGlyAlaSerArgAsnProCy 464
QY 711 CTCGAAACCTACTGTGGACTTTATCTCTGAGTCAGAACCAAGAGTGAAGCAGTGGGTAG 770
Db 464 sAspGluThrTyrCysGlyProAlaAlaGluSerGluLysGluThrLysAlaLeuAlaAs 484
QY 771 TTTCTTTGAGAGAAATATCAACAGATTAAGCATATCAGCATGATGATTCATCTACTATCCCA 830
Db 484 pPheIleArgAsnLysLeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGl 504
QY 831 GCATATAGTGTCTTCCATATTCTTATACACGAAGTAAAGCAAGACCATGAGGAACGTGC 890
Db 504 nMetMetIleTyrProTyrSerTyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAs 524
QY 891 TCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGTACAAAACTAGTAAATAATACCAGGTATAC 950
Db 524 nAlaLeuAlaLysAlaThrValLysGluLeu---AlaSerLeuHisGlyThrLysTyrTh 543
QY 951 ACATGGCCATGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACCATTCGATCTA 1010
Db 543 rTyrGlyProGlyAlaThrThrIleTyrProSerAlaGlyThrSerLysAspTrpAlaTy 563
QY 1011 TGATTTGGGCATCAAATATTCTGTTTACATC----- 1040
Db 563 rAspGlnGlyIleArgTyrSerPheThrPheGluLeuArgAspThrGlyArgTyrGlyPh 583
QY 1041 -----AAACCCACCTGTAGAGAAAGCTTTTCCCGCTGTC 1073
Db 583 eLeuLeuProGluSerGlnIleArgAlaThrCysGluGluThrPheLeuAlaIle 601
```

Search completed: January 1, 2004, 19:25:51

Job time : 68.8326 secs

Db 1 MetLysLeuCySerLeuAlaValLeuValProileValLeuPheCysGluGlnHisVal 20
Qy 61 TTCGGTTCAGAGTGGCCAGTTCTAGCTGCTCTTCTAGAACCTTAGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGAACTTACTACAACTATGAGATTCTTCTGGCAGCGGTAAACAGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrThrTyGluLeuValLeuTrpGlnProValThrAlaAsp 60
Qy 181 CTTATTGTGAAGAAAAAAGTCCATTTTTTTGTAATGCATCTGATGTCGCAATGTG 240
Db 61 LeuileValLysLysGlnValHisPhePheValAsnAlaSerAspValAspVal 80
Qy 241 AAAGCCATTAAATGTAGCGGAAATCCATCGAGTCTCTGCGCAGAGCTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyileProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGATTTCACACGACACAGTCAAGCCCGAGCTCCGATCTACTAT 360
Db 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyTy 120
Qy 361 GAACAGTATCACTCACTAAATGAATCTATTCTTTGGATAGAAATTTATACTGAGGSCAT 420
Db 121 GluGlnTyHisSerLeuAsnGluileTy-SerTrpIleGluPheIleThrGluArgHis 140
Qy 421 CCTGATATGTTCAAAAATCCACATTTGGATCTCATTTGAGAAGTACCCACTTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeuTyVal 160
Qy 481 TTTAAGGTTTCTGGAAGAACAAACAGCCAAAATGCCATATGATGATTGCTGGAAATC 540
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Qy 541 CATGCCAGAGATGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTAGGCCAT- 591
Db 181 HisAlaArgGlnTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
Qy 591 ----- 591
Db 201 PheTyGlyIleIleGlyGlnTyThrAsnLeuLeuArgLeuValAspPheTyValMet 220
Qy 592 -----AATCGAATGTGGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyAspTy-SerTrpLysLysAsnArgMetTrpArgLys 240
Qy 610 AACGGTTCTTTCTATGCGAACAACTATTGTCATCGGAACAGACCTGGAATAGCACTTTGTC 669
Db 241 AsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
Qy 670 TCCAAACACTGGTGTAGGAAGGTGCATCCAGTTCCTCTGCTCGGAAACCTACTGTGGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyCysGly 280
Qy 730 CTTTATCTCTGAGTACAGAACAGAGTGAAGTCAGTGGCTAGTTCTTTGAGAAGAAATATC 789
Db 281 LeuTyProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 300
Qy 790 AACAGATTAAGCATACATCAGCATGATTCATCTACTCCAGCATATAGTGTTCATAT 849
Db 301 AsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleValPheProTy 320
Qy 850 TCCTATACAGAACTAAAGCAAGACCATGAGAACTGCTCTAGTAGGCAAGTGAAGCA 909
Db 321 SerTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
Qy 910 GTTTCGTCTATGACAAAACCTAGTAAATAACACAGGTATACATGCGCATGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisGlyHisGlySerGlu 360
Qy 970 ACCTTATACCTAGTCTCTGGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATAT 1029
Db 361 ThrLeuTyLeuAlaProGlyGlyGlyAspAspTrpIleTyAspLeuGlyIleLysTy 380

Qy 1030 TCCTTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyGlyPheLeuLeuProGluArgTy 400
Qy 1038 ATCAAAACCCACCTGTGAGAAAGCTTTTCCGCTGCTCTCTAAA 1079
Db 401 IleLysProThrCysArgGluAlaPheAlaValSerLys 414
RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C:Accession: A32129; SI7543
R:Clauuser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A:Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compar:
A:Reference number: A92693; MUID:89034324; PMID:3182872
A:Accession: A32129
A:Molecule type: DNA
A:Residues: 1-415 <CUA>
A:Cross-references: GB:M23959; GB:J04041; NID:G203293; PIDN:AAA40872.1; PID:G203295
R:Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A:Title: Distribution of manganese in rat pancreas and identification of its primary bin
A:Reference number: SI7543; MUID:91378950; PMID:1898371
A:Accession: SI7543
A:Molecule type: protein
A:Residues: 109-116,'X',118-130 <KOD>
C:Genetics:
A:Gene: CPB
A:introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:354,376/Active site: Tyr, Glu #status predicted
Alignment Scores:
Pred. No.: 2,35e-52 Length: 415
Score: 692.00 Matches: 150
Percent Similarity: 55.09% Conservative: 61
Best Local Similarity: 39.16% Mismatches: 126
Query Match: 35.82% Indels: 46
DB: 2 Gaps: 7
US-09-980-881a-1_COPY_18_1097 (1-1080) x A32129 (1-415)
Qy 1 ATGAAGCTTTTGCAGCCTTGACGCTCTTGTACCCATTTCTCTTCTGTGAGCAGCATGTC 60
Db 1 MetLeuLeuLeuLeuAlaLeuValSerValAlaLeuAlaHisAlaSerGluGluHis- 19
Qy 61 TTCGGTTCAGAGTGGCCAGTTCTAGCTGCTCTTAGAACCTCTAGCAAGTTCAA 120
Db 20 -----PheAspGlyAsnArgValTyArgValSerValHisGlyGluAspHisValAsn 37
Qy 121 GTTCTACAGAACTTACTACAACTATGAGATTGTTCTCTGCGCAGCGGTAAACAGCTGAC 180
Db 38 LeuileGlnGluLeuAlaAsnThrLysGluileAspPheTrpLysProAspSerAlaThr 57
Qy 181 CTTATTGTGAAGAAAAAAGTCCATTTTTTTGTAATGCATCTGATGTCGCAATGTG 240
Db 58 GlnValLysProLeuThrThrValAspPheHisValLysAlaGluAspValAlaAspVal 77
Qy 241 AAAGCCATTAAATGTAGCGGAAATTCATGAGTGTCTTGTGGCAGAGCTGGAAGAT 300
Db 78 GluAsnPheLeuGluGluAsnGluValHisTyGluValLeuIleSerAsnValArgAsn 97
Qy 301 CTTATTCAACAGCAGATTTCACACGACACAGTCAAGCCCGAGCTCCGATCTACTAT 360
Db 98 AlaLeuGluSerGlnPheAspSerHisThr-----ArgAlaSerGlyHisSerTy 114
Qy 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGGSCAT 420

US-09-980-881A-1_COPY_18_1097 (1-1080) x A42332 (1-416)

QY	7	CTTTGCAGCCTTGCACTCTTGTAACCAATGTTCTCTCTGTGAGCAGCATGCTTCGCG	66
Db	1	MetLeuAlaLeuValThrValAlaLeuAlaSerAlaHisGlyGlyGlu	20
QY	67	---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTCTCTAGAACCTCTAGGCAAGTTCAGTT	123
Db	21	HisPheGluGlyGlyValPheArgValAsnValGluAspGluAsnHisIleAsnIle	40
QY	124	CTACAGAACTTACTACACATATGAGATGTTCTCTGGCAGCCGGTAACAGCTGACCTT	183
Db	41	IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln	60
QY	184	ATTGTGAAGAAAAACAAGTCCATTTTTTGTAAATGCATCTCATCGACAATGTGAAA	243
Db	61	IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu	80
QY	244	GCCATTAAATGTGAGCGGAATTCATCAGGTGTTCTGTGGCAGACGTGGGAAGACTT	303
Db	81	AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal	100
QY	304	ATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCGAGCCCTCCGCATCGTACTATGA	363
Db	101	ValGluAlaGlnPheAspSerArgVal-----ArgAlaThrGlyHisSerTyrGlu	117
QY	364	CAGTATCACTCACTAAATGAAATCTATTCTGGATAGAATTTATTAATCGAGGCATCCT	423
Db	118	LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro	137
QY	424	GATATCTTACAAAAATCCACATTTGGATCCTCATTTTGAGAAGTACCCTCTATGTTTA	483
Db	138	AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu	157
QY	484	AGGTTTCTGGAAGAACACAAACAGCCAAAATGCCATATGATGATCCTGTGGAATCCAT	543
Db	158	LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis	176
QY	544	GCCAGAGAATGGATCTCTCTGCTTTCTGCTGTGTTTCATA-----	585
Db	177	AlaArgGluTrpLysSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr	196
QY	585	-----	585
Db	197	TyrGlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuPro	216
QY	586	-----GGCCAT-----AATCGAATGTGGAGAAAGAAC	612
Db	217	ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr	236
QY	613	CGTCTTCTTATCGCAACAATCATTCATCGGAACAGACCTGAATAGCAACTTGTCTCC	672
Db	237	ArgSerThrHisThrGlySer---SerIleGlyThrAspProAsnArgAsnPhe---Asp	254
QY	673	AAACACTGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGACATT	732
Db	255	AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro	274
QY	733	TATCCTGAGTCAGAACGAGAGGAGGAGGTGGCTAGTTTCTTGAGAGAAATATCAAC	792
Db	275	AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer	294
QY	793	CAGATTAAAGCATACATCAGCATGCAATTCATCTCCACAGATATAGTGTTCATATTC	852
Db	295	SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer	314
QY	853	TATACAGAAGTAAAAAGCAAAACCATCTCTCTAGTAGCCAGTGAAGCAGATT	912
Db	315	TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal	334
QY	913	CGTGCTATTGACAAAAACTAGTAAAAATACCAGGTATACATATGCGGCATCGCTCAGAAAC	972
Db	335	LysGluLeu---AlaSerLeuHisGlyTrpLysTyrThrTyrGlyProGlyAlaThrThr	353

```
Qy 973 TTATACCTAGCTCTCGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATATTTCG 1032
    ::::|||||
Db 354 IletyrProAlaAlaGlyGlySerAspTrpAlaTyrAspGlnGlyIleArgTyrSer 373
    ::::|||||
Qy 1033 TTTACATC----- 1040
    ::::|||||
Db 374 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 393
    ::::|||||
Qy 1041 AAACCCACCTGTAGAGAGCTTTTGGCGCTCTC 1073
    ::::|||||
Db 394 ArgAlaThrCysGluGluThrPheLeuAlaIle 404
    ::::|||||

RESULT 6
CPBOB
carboxypeptidase B (EC 3.4.17.2) - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 02-Aug-1994
C/Accession: A93797; A92150; A00912
R/Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A/Title: Amino-acid sequence of bovine carboxypeptidase B.
A/Reference number: A93797; UID:75217824; PMID:1057162
A/Accession: A93797
A/Molecule type: protein
A/Residues: 1-306 <TIT>
R/Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A/Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations of
A/Reference number: A92150; UID:74260705; PMID:4833744
A/Accession: A92150
A/Molecule type: protein
A/Residues: 31-93|131-181;263-265;292-306 <SCH>
R/Schmid, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A/Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A/Reference number: A92839; UID:76265065; PMID:957425
A/Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R/Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A/Title: Isolation and sequence of peptides at the active center of bovine carboxypeptid
A/Reference number: A92051; UID:70007159; PMID:5344132
A/Contents: annotation; active site
R/Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A/Title: Identification of a glutamic acid at the active center of bovine carboxypeptid
A/Reference number: A92119; UID:73061487; PMID:4565668
A/Contents: annotation; active site
C/Superfamily: carboxypeptidase
C/Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;63-76,135-159,149-163/Disulfide bonds: #status experimental
F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F;246,268/Active site: Tyr, Glu #status experimental

Alignment Scores:
Pred. No.: 1-05e-42 Length: 306
Score: 581.50 Matches: 120
Percent Similarity: 56.66% Conservative: 46
Best Local Similarity: 40.96% Mismatches: 70
Query Match: 30.10% Indels: 58
DB: 1 Gaps: 6

US-09-980-881A-1_COPY_18_1097 (1-1080) x CPBOB (1-306)

Qy 358 TATGACAGTATCATCTCACTAAATGAATTTCTTGGATAGAAATTTAACTGAGAGG 417
    :|||:|||||
Db 6 TyrGluIleTyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValAlaSerGlu 25
    :|||:|||||
Qy 418 CATCTGATATGCTTACAAAATCCATTCGATCCCTATTGAGATCCCACTCTAT 477
    :|||:|||||
Db 26 AsnProAspLeuIleSerArgSerAlaIleGlyThrThrPheLeuGlyAsnThrIleTyr 45
    :|||:|||||
Qy 478 GTTTTAAAGGTTTCTGGAAAGACAAACAGCCAAATAATGCCATATGATGTTGACTGTGGA 537
```

```
Db 46 LeuLeuIleVal---GlyLysProGlySerAsnLysProAlaValPheMetAspCysGly 64
    ::::|||||
Qy 538 ATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTTGTGGTTTCATA----- 585
    :|||:|||||
Db 65 PheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaVal 84
    :|||:|||||
Qy 585 ----- 585
    :|||:|||||
Db 85 ArgThrTyrGlyArgGluIleHisMetThrGluPheLeuAspLysLeuAspPheTyrVal 104
    :|||:|||||
Qy 586 -----GGCCAT-----AATCGAATGTGGAGA 606
    :|||:|||||
Db 105 LeuProValValAsnIleAspGlyTyrIleTyrThrTrpThrThrAsnArgMetTrpArg 124
    :|||:|||||
Qy 607 AAGAACCGTTCTTCTATGCGAACAATCATTCGATCGGAACACACACCTGAATAGCACTTT 666
    :|||:|||||
Db 125 LysThrArgSerThrArgAlaGlySerSerCysThrGlyThrAspLeuAsnArgAsnPhe 144
    :|||:|||||
Qy 667 GTCTCCAAACACCTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 726
    :|||:|||||
Db 145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163
    :|||:|||||
Qy 727 GGAGTTTATCTCTGAGTCCAGAACCAAGTGAAGGCGATGCTAGTTCTTTGAGAAGAAAT 786
    :|||:|||||
Db 164 GlySerAlaAlaGluSerGluLysGluSerLysAlaValAlaAspPheIleArgAsnHis 183
    :|||:|||||
Qy 787 ATCAACGAGTTAAAGCATCATCAGTCATTCATATCCAGCATATAGTTGTTTCCA 846
    :|||:|||||
Db 184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetLeuTyrPro 203
    :|||:|||||
Qy 847 TATTCTATACACGAAGTAAAGCAAGACCATGAGGAACTGCTCTAGTAGCCAGTGAA 906
    :|||:|||||
Db 204 TyrSerTyrAspTyrLysLeuProLysAsnAsnValGluLeuAsnThrLeuAlaLysGly 223
    :|||:|||||
Qy 907 GCAGTTTGTGCTGCTATTGACAAAATAGTAAATAATACCAAGGTATATACATGCCCATGCTCA 966
    :|||:|||||
Db 224 AlaValLysLysLeu---AlaSerLeuHisGlyThrThrTyrSerTyrGlyProGlyAla 242
    :|||:|||||
Qy 967 GAAACCTTATACCTAGTCTCTGAGGTGGGAGCATGGATCATGATTTGGGCATCAAA 1026
    :|||:|||||
Db 243 ThrIleTyrProAlaSerGlyGlySerAspTrpAlaTyrAspGlnGlyIleLys 262
    :|||:|||||
Qy 1027 TATTGTTTATCATC----- 1040
    :|||:|||||
Db 263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282
    :|||:|||||
Qy 1041 -----AAACCCACCTGTAGAGAGCTTTTGGCGCTGTC 1073
    :|||:|||||
Db 283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIle 295
    :|||:|||||

RESULT 7
A38395
mast cell carboxypeptidase (EC 3.4.17.2) - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C/Accession: A38395; A33118
R/Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A/Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme activ
A/Reference number: A38395; UID:91105153; PMID:1988052
A/Accession: A38395
A/Status: preliminary
A/Molecule type: protein
R/Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A/Reference number: A33118
A/Accession: A33118
A/Status: preliminary
A/Molecule type: protein
R/Residues: 2-309 <CO2>
C/Superfamily: carboxypeptidase
```



```
Db 211 LeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrValPheSerGlnThrThr 230
|||||
Qy 592 AATCAATGTGGAGAAACCGTCTTCTTATCGCAACATCAATGTCATCGGAACAGAC 651
|||||
Db 231 AsnArgMetTrpArgLysThrArgSerLysArgSerGlySerGlyCysValGlyValAsp 250
|||||
Qy 652 CTGAATAGCACTTGTCTCCAAACACTGGTGTGAGGAGGTGCATCCAGTCTCTCATGC 711
|||||
Db 251 ProAsnArgMetTrp---AspAlaAsnPheGlyGlyProGlyAlaSerSerSerProCys 269
|||||
Qy 712 TCGAAACCTACTCTGGACTTTATCTCGAGTCAGAACACAGAGTGAAGCGAGTGGCTAGT 771
|||||
Db 270 SerAspSerTyrHisGlyProLysProAsnSerGluValGluValLysSerIleValAsp 289
|||||
Qy 772 TTCTTGAGAGAAATATACACAGATTAAAGCATATACATACATGATGATGATTCATCCAG 831
|||||
Db 290 PheIleLysSerHis---GlyLysValLysAlaPheIleThrLeuHisSerTyrSerGln 308
|||||
Qy 832 CATATAGTGTTCATATCTCTATACAGAGTAAAGCAAGACCATGAGGACTGCTCT 891
|||||
Db 309 LeuLeuMetPheProTyrGlyTyrLysCysThrLysProAspPheAsnGluLeuAsp 328
|||||
Qy 892 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTACACAAAACCTAGTAAATAACAGGTATACA 951
|||||
Db 329 GluValAlaGlnLysAlaAlaGluAlaLeu---LysArgLeuHisGlyThrSerTyrLys 347
|||||
Qy 952 CATGGCAGTCTCAGAAACCTTATACATGCTCTCTGAGTGGGAGCATGATGATCTAT 1011
|||||
Db 348 ValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySerIleAspTrpAlaTyr 367
|||||
Qy 1012 GATTGGGCATCAATATCTGTTTACATCAAC----- 1044
|||||
Db 368 AspLeuGlyIleLysTyrSerPheAlaPheGluLeuArgAspThrAlaPheTyrGlyPhe 387
|||||
Qy 1045 ---CCACCTGTAGAGAGCTTTTCCCG 1068
|||||
Db 388 LeuLeuProAlaLysGlnIleLeuPro 396
|||||

RESULT 9
CPBOA
carboxypeptidase A (EC 3.4.17.1) precursor [validated] - bovine
N:Alternate names: procarboxypeptidase A activation peptide
C:Species: Bos primigenius taurus (cattle)
C>Date: 24-Apr-1994 #sequence_revision 30-Jun-1992 #text_change 05-May-2000
C:Accession: JN0126; J0440; A31406; S00059; A38834; A00910; A90355
R:Le Hueerou, I.; Guilleoteau, P.; Touillec, R.; Puigserver, A.; Wicker, C.
Biochem. Biophys. Res. Commun. 175, 110-116, 1991
A:Title: Cloning and nucleotide sequence of a bovine pancreatic procarboxypeptidase A
A:Reference number: JN0126; MUID:91151335; PMID:1998496
A:Accession: JN0126
A:Molecule type: mRNA
A:Residues: 1-419 <LEH>
A:Cross-references: GB:M61851; NID:g162788; PIDN:AAA30426.1; PID:g162789
R:Fogliizzo, E.; Bonicel, J.; Kerfelec, B.; Granon, S.; Chapus, C.
Biochim. Biophys. Acta 954, 183-188, 1988
A:Title: Primary structure of the activation peptide from bovine pancreatic procarboxypeptidase A
A:Reference number: J0440; MUID:88209583; PMID:3365436
A:Accession: J0440
A:Molecule type: protein
A:Residues: 17-110 <FOG>
R:Wade, R.D.; Hass, G.M.; Kumar, S.; Walsh, K.A.; Neurath, H.
Biochimie 70, 1137-1142, 1988
A:Title: The amino acid sequence of the activation peptide of bovine pro-carboxypeptidase A
A:Reference number: A31406; MUID:89150306; PMID:3147705
A:Accession: A31406
A:Molecule type: protein
A:Residues: 17-110 <WAD>
R:Chapus, C.; Kerfelec, B.; Fogliizzo, E.; Bonicel, J.
Eur. J. Biochem. 166, 379-385, 1987
A:Title: Further studies on the activation of bovine pancreatic procarboxypeptidase A by
A:Reference number: S00059; MUID:87275921; PMID:3609014
A:Accession: S00059
```

```
A:Molecule type: protein
A:Residues: 17-69 <CHA>
R:Bradshaw, R.A.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 63, 1389-1394, 1969
A:Title: The amino acid sequence of bovine carboxypeptidase A.
A:Reference number: A38834; MUID:70050818; PMID:5260942
A:Accession: A38834
A:Molecule type: protein
A:Residues: 111-137, 'Q', 139-140, 'E', 142-417 <BRA>
R:Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 961-972, 1971
A:Title: Amino acid sequence of bovine carboxypeptidase A. Isolation and characterization
A:Reference number: A90360; MUID:71155185; PMID:4927805
A:Accession: A00910
A:Molecule type: protein
A:Residues: 214-411 <BR2>
A:Note: this is the final paper in a series
R:Petra, P.H.; Hermodson, M.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 4023-4025, 1971
A:Title: Characterization of bovine carboxypeptidase A(Allan).
A:Reference number: A90355; MUID:72138789; PMID:5143102
A:Accession: A90355
A:Molecule type: protein
A:Residues: 111-142 <PET>
R:Petra, P.H.; Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 8, 2762-2768, 1969
A:Title: Identification of the amino acid replacements characterizing the allotropic form
A:Reference number: A90556; MUID:69283620; PMID:5817619
A:Contents: annotation; allelic variant
A:Note: an allelic variant having 179-Val, 228-Glu, and 305-Leu occurs with nearly equal
R:Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A:Reference number: A50963; PDB:4CPA
A:Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137, 'Q', 139-141
R:Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 169, 367-387, 1983
A:Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A:Reference number: A92902; MUID:83294519; PMID:6887246
A:Contents: annotation; X-ray crystallography, 1.54 angstroms
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status experimental <ACP>
F:111-419/Product: carboxypeptidase A #status experimental <MAT>
F:179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F:248-271/Disulfide bonds: #status experimental
F:358,380/Active site: Tyr, Glu #status experimental

Alignment Scores:
Pred. No.: 1,68e-38 Length: 419
Score: 533.50 Matches: 134
Percent Similarity: 49.87% Conservative: 61
Best Local Similarity: 34.27% Mismatches: 133
Query Match: 27.61% Indels: 63
DB: 1 Gaps: 11

US-09-980-881A-1_COPY_18_1097 (1-1080) x CPBOA (1-419)
Qy 7 CTTTGCAGCCTTCAGTCTTGATCCCATGTTCTCTCTGTGAGCAGCATGTCTTCGCG 66
Db 4 LeuLeuIleLeuSerValLeuLeuGlyAlaAlaLeuGlyLysGluAsp----- 19
Qy 67 TTCAGAGTGGCAGTTCTAGCTGCTCTTCTAGAACCTCTTAGGACAGTTCAAGTTCTA 126
Db 20 PheValGlyHisGlnValLeuArgIleThrAlaAlaAspGluAlaGluValGlnThrVal 39
Qy 127 CAGAATCTTACTACACATATGAGTTGTCTC-----TGG-----CAGCCG 168
Db 40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProGlyGlnPro 59
Qy 169 GTAAACAGCTGACCTTATTGTGGAAGAAAAACAAGTCCATTTTTTTGTAATATGCATCTGAT 228
Db 60 -----GlySerProIleAspValArgValProPheProSer 71
```

QY	229	GTGCGA	CAATGTG	GAAGCC	CAATT	TAAATGTG	AGCGGA	ATTCCATG	CAGTGTCTTGTCTG	CGCA	288			
Db	72	LeuGln	AlaVal	ValPhe	LeuGlu	AlaHis	GlyIle	ArgTyr	ArgIle	MetIle	Glu	91		
QY	289	GACGTG	GAAGATCT	TATT	-----	CAACAG	CAGATT	TCCAA	CGACAG	CACTGAC	CCCC	339		
Db	92	AspVal	GlnSer	LeuLeu	AspGlu	GlnGlu	GlnMet	PheAla	SerGln	SerArg	Ala	111		
QY	340	CGAGCCT	CCGATCG	TAC	---TAT	GAA	CAGTAT	CAC	TAAATG	CAATCTAT	TCTTGG	396		
Db	112	ArgSer	ThrAsn	ThrPhe	AsnTyr	AlaThr	TyrHis	ThrLeu	AspGlu	IleTyr	AspPhe	131		
QY	397	ATAGA	TTTATA	CTGAG	AGCAT	CCTGAT	GTCTT	CAAAA	ATCCAT	TGGAT	CTCTCA	456		
Db	132	MetAsp	LeuLeu	AlaGlu	HisPro	GlnLeu	ValSer	LysLeu	GlnIle	GlyArg	Ser	151		
QY	457	TTTCAGA	AGTAGT	CCCACT	CTATG	TTTTAA	AGGTTTCT	GGAAGA	AGACAA	CAACGCCA	AAAAAT	516		
Db	152	TyrGlu	GlyArg	ProIle	TyrVal	LeuLys	PheSer	---Thr	GlyGly	SerAsn	ArgPro	170		
QY	517	GCCATAT	GGAATG	ACTGTG	GGAAT	CCATCC	ACGAGAA	TGGAT	CTCTCT	CTCTTCTG	CTTGG	576		
Db	171	AlaIle	TrpIle	AspLeu	GlyIle	HisSer	ArgGlu	TrpIle	ThrGln	AlaThr	GlyVal	190		
QY	577	TGGTT	TCATA	-----	-----	-----	-----	-----	-----	-----	-----	585		
Db	191	TrpPhe	AlaTyr	LysPhe	ThrGlu	AspTyr	GlyGln	AspPro	SerPhe	ThrAla	IleLeu	210		
QY	585	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	585		
Db	211	AspSer	MetAsp	IlePhe	LeuGlu	IleVal	ThrAsn	ProAsp	GlyPhe	AlaPhe	ThrHis	230		
QY	586	GGCCAT	AATCGA	ATGTG	GAGAA	AGACCG	TCTTCT	TATG	GCAACA	ATCAAT	TGCA	CGCA	645	
Db	231	SerGln	AsnArg	LeuTrp	ArgLys	ThrArg	SerVal	ThrSer	SerSer	LeuCys	ValGly	250		
QY	646	ACAGAC	CTGA	TAGCA	ATTG	CTCTC	CAACAC	ACTGGT	GTGAGG	AAAGTGC	ATCCAG	TCC	705	
Db	251	ValAsp	AlaAla	AsnArg	AsnTrp	---Asp	AlaGly	PheGly	LysAla	GlyAla	SerSer	Ser	269	
QY	706	TCATG	CTCGG	AAACCT	ACTG	TGGAGT	TTTAT	CCTG	AGTCAG	ACACG	AGAGTGA	AGCAG	CTG	765
Db	270	ProCys	SerGlu	ThrTyr	HisGly	LysTyr	AlaAsn	SerGlu	ValGlu	ValLys	SerIle	289		
QY	766	GCTAGT	TTCTTG	AGAAGA	ATAT	CAAC	CAGATTA	AAAGCAT	ATACAT	CAGCAT	GCAAT	TCAT	CA	825
Db	290	ValAsp	PheVal	LysAsp	HisGly	Asn---Phe	LysAla	PheLeu	SerIle	HisSer	Tyr	308		
QY	826	TCCAG	CATATAG	TGTTT	CCATAT	TCTT	CATAC	AGAA	TAAAG	CAAGAC	CATGAG	AA	885	
Db	309	SerGln	LeuLeu	LeuTyr	ProTyr	GlyTyr	ThrThr	GlnSer	IlePro	AspLys	ThrGlu	328		
QY	886	CTGTCT	CTAGT	PAGC	AGTGA	GCAGT	TCGTCT	ATTG	ACAAA	CTAGT	ATAAAA	TACC	AGG	945
Db	329	LeuAsn	GlnVal	AlaLys	SerAla	ValGlu	AlaLeu	---Lys	SerLeu	TyrGly	ThrSer	347		
QY	946	TATAC	ATAG	CCCG	CTG	CTCAG	AACCTT	TATAC	TCTAG	CTCTG	GAGTGGG	GAC	ATTGG	1005
Db	348	TyrLys	TyrGly	SerIle	IleThr	ThrIle	TyrGln	AlaSer	GlyGly	SerIle	AspTrp	367		
QY	1006	ATCTAT	GATTTG	GGCATC	CAATA	TATTCG	TTTACA	1038						
Db	368	SerTyr	AsnGln	GlyIle	LysTyr	SerPhe	Thr	378						

A>Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of
A:Reference number: A00911; MUID:82105986; PMID:6275388
A:Accession: A00911
A:Molecule type: mRNA
A:Residues: 1-419 <QUI>
A:Cross-references: GB:J00713; MID:g203364; PIDN:AAA40893.1; PID:g203365
R:Clausner, B.; Gardell, S.-J.; Craik, C.S.; Macdonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A>Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparison
A:Reference number: A92693; MUID:89034324; PMID:3182872
A:Accession: B32129
A:Molecule type: DNA
A:Residues: 1-260, 'F', 262, 'M', 264-346, 'K', 348-419 <CLA>
A:Gene: CPAL
A:Genetics: C
A:Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A:Note: the authors translated the codon GTC for residue 196 as Ile
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status predicted <PRO>
F:111-419/Product: carboxypeptidase A #status predicted <MAT>
F:179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F:248-271/Disulfide bonds: #status predicted
F:358,380/Active site: Tyr, Glu #status predicted

[illegible]

US-09-980-881A-1 COPY 18 1097 (1-1080) x CPRTA (1-419)

Qy	4	AAGCTTTGCAGCCTTGCAGTCTCTTGTATACCAATGTTCTCTCTGTGAGCAGCATGCTTTC	63
Db	3	ArgLeuLeuLeuLeuSerLeuLeuLeuAlaValCysGlyAsnGluAsn-----	19
Qy	64	CGGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTTCTCTAGAACCTCTAGGCAGTTCAGATT	123
Db	20	---PheValGlyHisGlnValLeuArgIleSerAlaAspGluAlaGlnValGlnIys	38
Qy	124	CTACGAATCTTACTACAACATATGAGATTGTTCTC-----TGGCAGCCGCTAACAGCT	177
Db	39	ValIysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgAspAlaAlaArg	58
Qy	178	GACCTTATGTGAAGAAAAACAAGTCCATTTTTTTGTGAATGCATCTGATGTGCACAA	237
Db	59	AlaGlyIleProIleAspValArgValProPhe-----ProSerIleGlnSer	74
Qy	238	GTGAAGCCCATTTAAATGTGAGCGGAATCCATGCAGTGTCTGTCTGCAGACGTGGA	297
Db	75	ValIysAlaPheLeuGluIuTyHisGlyIleSerTyrgluIleMetIleGluAspValGln	94
Qy	298	GATCTTATT-----CAACACAGATTTCCAACACACAGTCAGCCCCCGAGCCTCC	348
Db	95	LeuLeuLeuAspGluGluIysGlnGlnMetSerAlaPheGlnAlaArgAlaLeuSerThr	114
Qy	349	GCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA	405
Db	115	AspSerPheAsnTyralaThrTyHisThrLeuAspGluIleTyrgluPheMetAspLeu	134
Qy	406	ATACTGAGAGGCATCTGTATATGCTTACAAAAATCCACATTTGATCTCATTTGAGAG	465
Db	135	LeuValAlaGluHisProGlnLeuValSerIysIleGlnIleGlyAsnThrPheGluGly	154
Qy	466	TACCCACTCATGTTTTAAAGTTTCTCGAAAGACAAACAGCCAAAAATGCCCATGG	525
Db	155	ArgProIleHisValLeuIysPheSer---ThrGlyGlyThrAsnArgProAlaIleTrp	173
Qy	526	ATTGACTGTGGAATTCATGTCGAGAGAATGGATCTCTCTGCTTTCTGTGTGGTTCATA	585

```
Db 174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193
Qy 586 GGCCAT----- 591
Db 194 LysLysValThrLysAspTyrGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213
Qy 592 -----AAT 594
Db 214 AspilePheLeuGluIleValThrAsnProAspGlyPheAlaTyrThrHisLysThrAsn 233
Qy 595 CGAATGTGGAGAAAGAACCGTTCTTTCTATCGCAACAATCATTCATCGGAAACAGACCTG 654
Db 234 ArgMetTrpArgLysThrArgSerHisThrGlnGlySerLeuGlyValGlyValAspPro 253
Qy 655 AATAGCACTTGTCTCCAAACATCG-----TGTGAGGAGGTGCATCCAGT 702
Db 254 AsnArgAsn-----TrpAspAlaGlyLeuGlyLysAlaGlyAlaSerSer 268
Qy 703 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCTGAGTCAGAACCAAGAGTGAAGGCA 762
Db 269 AsnProCysSerGluThrTyrArgGlyLysPheProAsnSerGluValGluValLysSer 288
Qy 763 GTGCTAGTTCTTCTGAGAAGAAATATCAACAGATTAAAGCATACATCAGCATCATTCATCA 822
Db 289 IleValAspPheValThrSerHisGlyAsn---IleLysAlaPheIleSerIleHisSer 307
Qy 823 TACTCCAGCATATAGTGTTCATATCTTCATACAGAGTAAAGCAAGAACACCATGAG 882
Db 308 TyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrSerGluProAlaProAspGlnAla 327
Qy 883 GAACGTCTCTAGTAGCAGGAGGAGTTCGTCTATTGACAAACTAGTAAATAATACC 942
Db 328 GluLeuAspGlnLeuAlaLysSerAlaValThrAlaLeu---ThrSerLeuHisGlyThr 346
Qy 943 AGGTATACATACATGCCATGCTCGAAGAACCTTATACCTAGTCTCTGGAGGTGGGAGCAT 1002
Db 347 GluPheLysTyrGlySerIleLeuAspThrIleTyrGlnAlaSerGlySerThrIleAsp 366
Qy 1003 TGGATCTATGATTTGGGCAATCAATATTCGTTTACA 1038
Db 367 TrpThrTyrSerGlnGlyIleLysTyrSerPheThr 378

RESULT 11
A56171
Carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence revision 11-Aug-1995 #text_change 22-Jun-1999
R;Accession: A56171; S02809; S71395
R;Catalus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and characterization of human pancreatic procarboxypeptidase A2. cDNA
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:g790226; PIDN:AAA74425.1; PID:g790227
R;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: 'S',18-36,'N',38-43 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic procarboxypeptidase A1 a
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
```

```
A;Molecule type: mRNA
A;Residues: 1-301,'T',303-338,'SR',341-417 <LAB>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted
```

```
Alignment Scores:
Pred. No.: 2,85e-36 Length: 417
Score: 508.00 Matches: 128
Percent Similarity: 48.48% Conservative: 63
Best Local Similarity: 32.49% Mismatches: 137
Query Match: 26.29% Indels: 66
DB: 2 Gaps: 11
```

```
US-09-980-881A-1_COPY_18_1097 (1-1080) x A56171 (1-417)
Qy 55 CATGTCTTC-----GGTTCCAGAGTGGCCAGTTCTAGCTGCTCTTCTAGAAC 105
Db 13 HisIleTyrCysLeuGluThrPheValGlyAspGlnValLeuGluIleValProSerAsn 32
Qy 106 TCTAGGCAAGTTCAAGATTTCTACAGAATCTTACTACACATATGAGATTTGTTCTC 159
Db 33 GluGluGlnIleLysAsnLeuLeuGluAlaGlnGluHisLeuGlnLeuAspPhe 52
Qy 160 TGGAGCCGGTACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTGTGTAAT 219
Db 53 TrpLysSerProThr-----ThrProGlyGluThrAlaHisValArgValPro 68
Qy 220 GCATCTCATGCTCGACAATGTGAAGCCCATTTAAATGTGAGCGAATTTCCACAGCAG 279
Db 69 PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaTyrSerile 88
Qy 280 TTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGCATTTTCCACAGCAGCAGCC 339
Db 89 MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg 108
Qy 340 CGAGCCTCCGCATCG-----TACTATGAACAGTATCAGTCACTAAATGAATCAT 390
Db 109 ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaTyrHisThrLeuGluGluIleSer 128
Qy 391 TCTTGGATAGAATTATTAAGGATCTCTGATATGCTTACAAAAATCCACATTGGA 450
Db 129 GlnGluMetAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly 148
Qy 451 TCCTCATTTGAGAGTACCACCTCTATGTTTAAAGGTTTCT---GGAAAAGAACAAACA 507
Db 149 SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyGlyAsp 166
Qy 508 GCCAAAAATGCCATATGATTGATGTGGAATCCATGCCAGAGATGATCTCTCTCGCT 567
Db 167 ---LysProAlaIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla 185
Qy 568 TTCTGCTTGGTTCATAGC----- 588
Db 186 ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205
Qy 588 ----- 588
Db 206 SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal 225
Qy 589 -----CATATCGAATGTGGAGAAAGAACCGTTCTTCTTATGCGAACATCAT 636
Db 226 PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLysValSerGlySerLeu 245
```

Qy 637 TGCATCGGACACACCTGATGCAACTTTGTCTCCAAACACTGGTGTGAGGAGGTGCA 696
Db 246 CysValGlyValAspProAlaArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla 264
Qy 697 TCCAGTTCTCCATGCTCGGAAACCTACTGTGGACTTTTCTCCTGAGTCAGAACACAGAGTG 756
Db 265 SerSerAsnProCysSerAspSerTyRHiGlyProSerAlaAsnSerGluValGluVal 284
Qy 757 AAGCAGTGGCTAGTTCTTGTGAGAGAAATATCAACAGATTAAAGCATATCATCAGCATG 816
Db 285 LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleIleLeu 303
Qy 817 CATTATACATCCAGCATATAGTTTCCATATTCCTATACAGAGTAAGCAAGAC 876
Db 304 HisSerTyRSerGlnLeuLeuMetPheProGlyTyRLeuLysCysThrLysLeuAspAsp 323
Qy 877 CATGAGCAATGCTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAACTAGTAAA 936
Db 324 PheAspGluLeuSerGluValAlaGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis 342
Qy 937 AATACAGGTATACATGCGCATGCGCTCGAAGAACCTTATACCTAGTCTCTGGAGTGGG 996
Db 343 GlyThrLysTyRLeuValGlyProIleCysSerValIleTyRAlaSerGlyGlySer 362
Qy 997 GACGATTGGATCTATGATTGGGATCAATATTCGTTTACATCAAC----- 1044
Db 363 IleAspTrpSerTyRAspTyRLeuGlyIleLysTyRSerPheAlaPheGluLeuArgAspThr 382
Qy 1045 -----CCACCTGTAGAGAGCTTTTTCGCG 1068
Db 383 GlyArgTyRLeuLeuLeuProAlaArgGlnIleLeuPro 396

RESULT 12
S29127
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N;Alternate names: pancreatic carboxypeptidase A1
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence revision 19-Jan-1996 #text change 18-Jun-1999
C;Accession: S29127; S08253; S02810; S71394; S02811
R;Cataus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver
Biochem. J. 287, 299-303, 1992
A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.
A;Reference number: S29127; MUID:93038569; PMID:1417781
A;Accession: S29127
A;Molecule type: mRNA
A;Residues: 1-419 <CAT>
A;Cross-references: EMBL:X67318; NID:g35329; PIDN:CAA47732.1; PID:g35330
R;Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A;Title: Human carboxypeptidase A identifies a BgIII RFLP and maps to 7q31-qter.
A;Reference number: A34205; MUID:90196012; PMID:1969228
A;Accession: A34205
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 330-396 <STE>
A;Note: the authors translated the codon CTG for residue 391 as Val
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxypeptidase A1.
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08253
A;Molecule type: protein
A;Residues: 17-43; 'XXX', 114-135 <MOU>
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase A.
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A;Molecule type: protein
A;Residues: 17-42 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic procarboxypeptidase A1 a

A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71394
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-419 <LAE>
C;Genetics:
A;Gene: GDB:CPA1; CPA
A;Cross-references: GDB:120597; OMIM:114850
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.: 1.18e-34 Length: 419
Score: 489.50 Matches: 131
Percent Similarity: 47.63% Conservative: 60
Best Local Similarity: 32.67% Mismatches: 127
Query Match: 25.34% Indels: 83
DB: 1 Gaps: 13

US-09-980-881A-1_COPY_18_1097 (1-1080) x S29127 (1-419)

Qy 7 CTTTGCGAGCTTCAGTCCTTGTACCATTTGTTCTCTCTGTGAGCAGCATGTTCTCGCG 66
Db 4 LeuLeuValLeuSerValLeuLeuGlyAlaValPheGlyLysGluAsp----- 19
Qy 67 TTCAGAGTGGCCCAAGTTCTAGCTGTCTTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db 20 PheValGlyHisGlnValLeuArgIleSerValAlaAspGluAlaGlnValLysVal 39
Qy 127 CAGAACTTACTACACATATGAGATTGTTCTC-----TGGCAG----- 165
Db 40 LysGluLeuLeuAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProAlaHisPro 59
Qy 166 -----CCGTAACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTGTAAAT 219
Db 60 GlySerProIleAspVal-----ArgValProPhe----- 69
Qy 220 GCATCTGATGTCACAAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGTCAGTGTG 279
Db 70 ---ProSerIleGlnAlaValLysIlePheLeuGluSerHisGlyIleSerTyRLeuThr 88
Qy 280 TTGCTGGCAGACGTGGAAGATCTTATT-----CAACAGCAGATT----- 318
Db 89 MetIleGluAspValGlnSerLeuLeuAspGluGlnGluGlnMetPheAlaPheArg 108
Qy 319 -----TCCACGACAGCAGTCAGCCCGCCGCGCTCCGATCGTACTATGAACAG 366
Db 109 SerArgAlaArgSerThrAspThrPheAsn-----TyAlaThr 121
Qy 367 TATCAGCTCACTAAATGAATCTATTCTTGGATAGATTATTAAGTACGAGCAGCTTCGAT 426
Db 122 TyrHisThrLeuGluGluIleTyRAspPheLeuAspLeuValAlaGluAsnProHis 141
Qy 427 ATGCTTACAAAAATCCACATTTGGATCTCTATTGAGAAAGTACCCACTCTATGTTTTAAAG 486
Db 142 LeuValSerLysIleGlnIleGlyAsnThrTyRGluclyArgProIleTyRValLeuLys 161
Qy 487 GTTCTGGAAGAACACCAACGACCAAAATGCCATATGATTTGATTTGATTCGATCCATGCC 546
Db 162 PheSer---ThrGlyGlySerLysArgProAlaIleTrpIleAspThrGlyIleHiser 180
Qy 547 AGAATGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 585
Db 181 ArgGluTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThrGlnAspTyR 200
Qy 585 ----- 585

QY	907	GCAGTTTGGTCTATTGACAAACCTAGTAAATAATACCAAGGTATACACATGATGCCATGGCTCA	966
Db	368	AlaThrLysAlaLeuTyrGluLeu---TyrGlyThrLysTyrGlnValGlyThrGlyAla	386
QY	967	GAACACCTTATACCTAGCTCTCGGAGGTGGGACCATGGATCTCTAT---GATTTGGGCATC	1023
Db	387	AspLeuMetTyrGluAlaSerGlyGlySerHisAspTrpAlaLysGlyGlnLeuLysVal	406
QY	1024	AAATATTTCGTTT	1035
Db	407	ProTyrAlaTyr	410
RESULT 14			
T20507			
Hypothetical protein F02D8.4 - Caenorhabditis elegans			
C;Species: Caenorhabditis elegans			
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000			
C;Accession: T20507; T24289			
R;Matthews, L.			
submitted to the EMBL Data Library, August 1996			
A;Reference number: Z19284			
A;Accession: T20507			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: DNA			
A;Residues: 1-323 <WTL>			
A;Cross-references: EMBL:Z78411; PIDN:CA801647.1; GSPDB:GN00023; CESP:F02D8.4			
R;Wild, A.			
submitted to the EMBL Data Library, August 1996			
A;Reference number: Z19869			
A;Accession: T24289			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: DNA			
A;Residues: 1-323 <WLT>			
A;Cross-references: EMBL:Z78413; PIDN:CA801666.1; GSPDB:GN00023; CESP:F02D8.4			
A;Experimental source: clone T01C3			
C;Genetics:			
A;Gene: CESP:F02D8.4			
A;Map position: 5			
A;Introns: 33/3; 67/3; 201/3; 228/1; 257/2			
C;Superfamily: carboxypeptidase			
Alignment Scores:			
Pred. No.:	1-42e-26	Length:	323
Score:	397.00	Matches:	88
Percent Similarity:	53.01%	Conservative:	53
Best Local Similarity:	33.08%	Mismatches:	81
Query Match:	20.55%	Indels:	44
DB:	2	Gaps:	7
US-09-980-881A-1_COPY_18_1097 (1-1080) x T20507 (1-323)			
QY	394	TGGATGAAATTTTAACTGAGAGCAFCCTGATATGCTTACAAAATCCACATTGGATCC	453
Db	3	PhelLeuAsnSerLeuAlaGlnGlnTyrProAsnAspValLysLeuGlnAenlleGlyAsn	22
QY	454	TCATTTGAGAAGTACCACCTCTATGTTTAAAGTTTCTGAAAAGAACAAACAGCCAAA	513
Db	23	SerTyrGluGlyArgSerIleThrAlaValArgileAla---AspAspGlySerSerLys	41
QY	514	AATGCCATATGGATTGACTGGGAATCCATGCCAGAGAATGGATCTCTCTGCTTTTCG	573
Db	42	ProIleValTrpIleAspAlaGlyIleHisAlaArgLysTrpIleSerTyrAsnValAla	61
QY	574	TTGTGGTTTCATA	585
Db	62	LeuTyrLeuIleTyrThrIleValSerGlnProAlaTyrArgAsnLeuLeuAspSerVal	81
QY	586	-----GCCATAAT	594
Db	82	GlnLeuValValProAsnThrAsnProAspGlyTyrGluTyrSerArgThrAsnAsp	101

Qy	595	CGAATGGAGAAAGACCGTCTCTTCATATGCGAACAATCATTCGATCGGACAGACCTG	554
Db	102	ArgwecTrpArguylsThrArgSerArgPheThrAsnSerArgCysalagAlaAspAla	121
Qy	655	AATAGCAACTTTCTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATCTCG	714
Db	122	AsnArgAsnTyr---PropheTyrTrpGlyThrGlnGlyValSerHisserGlnCysSer	140
Qy	715	GAACACTTACTGTGGACTTTATCTCGAGTCAGAACACGAGAGTGAAGCAGTGGCTAGTTC	774
Db	141	GlulPheCysGlySerArgProGlnSerGluProGluValLeuAlaLeuThrAsnAla	160
Qy	775	TTGAGAGAATAATCAACGAGATTAAGCATATCATCAGCATGCATTCATCTCCACGAT	834
Db	161	IleIleArgAspGluGluArgilelylsGlyTyrIleAlaLeuHisserTyrGlyGlnGlu	180
Qy	835	ATAGTGTTCATATCCCTATACA---CGAAGTAAAGCAAGACCATCAGGAACTGTCT	891
Db	181	IleuTyrProTrpGlyHisThrGlnArgThrTyrProThrAspValGlnAspLeuLeu	200
Qy	892	CTTAGTACC-----AGTGAAGCAGTTCGTGCTATTTCACAAAACTAGTAAAAATACC	942
Db	201	GlnValGlyArgAlaMetAlaSerAlaIleArgAlaValAsn-----AsnThr	216
Qy	943	AGGTATACATCGGCATGGCTCGACAAACCTTATACCTAGTCTCGAGTGGGGACGAT	1002
Db	217	AspTyrThrValValAsnSerGlyAspGlyLeuTyrProAlaAlaGlyAlaSerAspAsp	236
Qy	1003	TGGCATATGATTGGGCATCAATATTTCGTTTACATCAAC---CCACTCTAGAGAAG	1059
Db	237	TrpAlaIlySerArgGlyIleuTyrSerTyrThrIleGluLeuSerProfileAspAsp	256
Qy	1060	CTTTTGGCGGTCTCTTA 1077	
Db	257	PheThrGlyPheSerLeu 262	
RESULT 15			
T33526			
hypothetical protein T06A4.3 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #ext_change 29-Oct-1999			
C:Accession: T33526			
R:Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.			
submitted to the EMBL Data Library, October 1998			
A:Description: The sequence of C. elegans cosmid T06A4.			
A:Reference number: Z21364			
A:Accession: T33526			
A>Status: preliminary; translated from GB/EMBL/DDBJ			
A:Molecule type: DNA			
A:Residues: 1-667 <WU>			
A:Cross-references: EMBL:AF098994; PIDN:AAC67474.1; GSPDB:GN00019; CESP:T06A4.3			
A:Experimental source: strain Bristol N2; clone T06A4			
C:Genetics:			
A:Gene: CESP:T06A4.3			
A:Map position: 1			
A:Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 5			
Alignment Scores:			
Pred. No.:		2,01e-26	Length: 667
Score:		395.50	Matches: 121
Percent Similarity:		41.91%	Conservative: 63
Best Local Similarity:		27.56%	Mismatches: 138
Query Match:		20.47%	Indels: 117
DB:		2	Gaps: 15
US-09-980-881A-1_COPY18_1097 (1-1080) x T33526 (1-667)			
Qy	34	ATGTGTTCTCTTCGTGAGCAGCATGTC---TTTCGGCTTCACAGTGGCCAGTTCCTAGCT	90
Db	8	LeuIlePheLeuCysPheSerAlaValSerLeuSerThrGlnSerPheAlaValTyrArg	27
Qy	91	GCCTCTTCTAGAACCTCTTAGCGAAGTTCAGTTCTTACAGATCTT-----ACTACA	141

Db 28 LeuLeuProLysSerGlnThrAspPheGlnAlaIleGlnArgLeuTyrLysAsnAlaThr 47
Qy 142 ACATATGAGATTGTTCTCGCAGCCGGTAACAGCT-----GACCTATT 186
Db 48 AspHisAspLeuAsnPheTrpLysThrGlyLysAspLysHisGlyPheTrpAspValMet 67
Qy 187 GTGAAGAAAAAACAAGTCCATTGTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
Db 68 ValAspMetLysAsnSerLysTyrPheLeuAsp----- 78
Qy 247 CATTTAAATGTGACGGAAATCCATGACGAGTCTTGTGTCGACAGCGTGAAGATCTTATT 306
Db 79 PheLeuGlnValAsnAspIleSerTyrIleLysThrIleAspValGluGlyLeuIle 98
Qy 307 CAACAGCAGATTTCCACGACACAGTCAGC-----CCCCGA----- 342
Db 99 ThrLysHisGluAsnAsnLysThrLeuSerAsnMetPheProArgLeuTrpAspAspSer 118
Qy 343 GCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAATCTATTCTTGATA 399
Db 119 SerSerAlaHisTyrAspPheHisThrTyrGlySerTyrGlnArgMetThrAspTrpMet 138
Qy 400 GAATTTATACTAGAGGCATCTGATATGCTTACAAAAATCCACATGGATCCTCATTT 459
Db 139 LysGlnLeuValValLysTyrProLysMetValGlnTyrIleSerIleGlyLysThr 158
Qy 460 GAGAAGTACCCACTCTATGTTTAAAGTTTCTGGAAGAAAGAACAAACAGCCAAAATGCC 519
Db 159 GluGlyArgAsnIleAspGlyValGluIleGlyLysSerArgThr---LysLysIle 177
Qy 520 ATATGGATTGACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCTTCTGCTGTTGG 579
Db 178 PheTrpIleAspGlyGlyIleHisAlaArgGlnTrpAlaAlaProHisThrAlaLeuPhe 197
Qy 580 TTCATA----- 585
Db 198 PheIleHisGlnValCysAspProAlaProSerGluLysSerValPheArgLeuValPro 217
Qy 585 ----- 585
Db 218 LeuGlyLeuHisLysCysTrpGlnArgGlnLysGlyLysLeuAspAsnLysTrpLeu 237
Qy 585 ----- 585
Db 238 ThrSerArgAlaAsnGluProGlyIleLysLysLeuAsnGluIleThrPheValVal 257
Qy 586 -----GCCATAAT 594
Db 258 ValProCysLeuAsnProAspGlyTyrGluPheThrArgSerSerThrAsnProHisVal 277
Qy 595 CGAATGTGGAGAAAGAACCGTTCT-----TTCTATGCCAAC 630
Db 278 ArgLeuTrpArgLysAsnArgSerLysMetGlnCysArgLysAspIleTrpGlyArgAsn 297
Qy 631 AATCATTGTCATCGGAACAGACCTGAATAGCACTTGTCTCAACACACTGGTGAGGAA 690
Db 298 ArgCysCysArgGlyValAspLeuAsnArgAsnPhe---AspPheHisPheArgGluSer 316
Qy 691 GGTGCATCCAGTCTCTCATGTCGGAACCTACTGTGACTTTTATCTCTGAGTCAGAACCA 750
Db 317 GlyThrSerAspAspProCysSerGluIleTyrGlnGlyProSerProPheSerGluPro 336
Qy 751 GAAGTGAAGGAGTGGCTAGTCTTG-----AGAAGAAATATCAACACAGATTAAAGCA 804
Db 337 GluAlaLysAlaValArgAspAlaLeuLeuSerGlnArgTyrLysGlyArgThrAspAla 356
Qy 805 TACATCAGCATGTCATCTCCAGCATATAGTGTTCATATTCCTATATACACAG--- 861
Db 357 TyrIleThrLeuHisThrTyrSerGlnIleTrpIleHisProTyrGlyHisLysLysAsp 376
Qy 862 AGTAAAGCAAGACCACTGTCTCTAGTAGCAGTGAAGCAGATTCTGCTATT 921
Db 377 AlaTyrProGlyAspIleLysAspLeuTyrGluValGlyLysLysAlaAlaGlnAlaLeu 396

Qy 922 GACAAAACTAGTAAAAATACAGGTATACACATGGCCATGGCTCAGAAAACCTTATACCTA 981
Db 397 ---LysArgValTyrGlyThrLysTyrValValGlySerGlyAlaAspThrLeuTyrPro 415
Qy 982 GCTCCTCGAGGTGGGACGATTCGATC---TATGATTTGGGCATCAATAATTCGTTT 1035
Db 416 AlaSerGlyGlySerGluAspTrpAlaLysHisGluAlaLysValLysPheValTyr 434

Search completed: January 1, 2004, 19:00:49
Job time : 39.1685 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:41:10 ; Search time 11.3984 Seconds
(without alignments)
8911.561 Million cell updates/sec

Title: US-09-980-881a-1_COPY_18_1097

Perfect score: 1932
Sequence: 1 atgaagcttgccgcttgc.....ttttgcgctgtctctataaa 1080

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2 1/USPTO spool/US09980881/runat 31122003 135748 24941/app query.fasta_1.2958
-DB=SwissProt 41 -OFWT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USFR=US09980881 @CGN 1.1 34 @runat 31122003 135748 24941 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DSLOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	692	35.8	415	1	CBPB_RAT P19223 rattus norv
2	641	33.2	417	1	CBPB_HUMAN P15086 homo sapien
3	638	33.0	417	1	CBPC_MOUSE P15089 mus musculu
4	635	32.9	417	1	CBPC_HUMAN P15088 homo sapien
5	629.5	32.6	401	1	CBPB_PIG P09955 sus scrofa
6	627.5	32.5	416	1	CBPB_CANFA P55261 canis fami
7	581.5	30.1	306	1	CBPB_BOVIN P20732 bos taurus
8	555.5	28.8	309	1	CBPC_RAT P19761 rattus norv
9	553	28.6	417	1	CBP2_RAT P19222 rattus norv
10	533.5	27.6	419	1	CBPA_BOVIN P00730 bos taurus
11	529.5	27.4	419	1	CBP1_RAT P00731 rattus norv
12	511	26.4	419	1	CBP2_HUMAN P48052 homo sapien
13	489.5	25.3	419	1	CBP1_HUMAN P15085 homo sapien
14	483	25.0	421	1	Q9ui42 homo sapien
15	404.5	20.9	304	1	CBP2_SIMVI P42788 simulium vi
16	386	20.0	433	1	CBPA_ANOGA O02350 anopheles g
17	369.5	19.1	303	1	CBPB_ASTFL P04069 astacus flu
18	299.5	15.5	430	1	YHT2_YEAST P38836 saccharomyc

19	277.5	14.4	424	1	CBPT_THEVU
20	271	14.0	451	1	CBPS_STRGR
21	251	13.0	434	1	CBPS_STRCP
22	118.5	6.1	707	1	ORC1_SCHPO
23	114.5	5.9	1380	1	CBPD_HUMAN
24	112.5	5.8	1377	1	CBPD_MOUSE
25	105	5.4	1255	1	PER2_HUMAN
26	103.5	5.4	722	1	CPXM_MOUSE
27	102	5.3	734	1	CPXM_HUMAN
28	100	5.2	376	1	YQGT_BACSU
29	99.5	5.2	445	1	TPH1_CHICK
30	98.5	5.1	488	1	TY3H_ANGAN
31	97	5.0	663	1	TAZI_SCHPO
32	95	4.9	444	1	TPH1_HUMAN
33	92	4.8	525	1	PUR9_BUCAI
34	91.5	4.7	465	1	STHA_ECO57
35	91.5	4.7	465	1	STHA_ECOLI
36	91.5	4.7	3588	1	SRF1_BACSU
37	91	4.7	297	1	Y103_HUMAN
38	91	4.7	444	1	TPH1_RABIT
39	91	4.7	982	1	ENV_SFV3L
40	90	4.7	1451	1	SPT6_YEAST
41	89.5	4.6	682	1	CJBA_BACUH
42	88.5	4.6	319	1	UN30_CAEEL
43	88.5	4.6	388	1	RFC5_CAEEL
44	88.5	4.6	466	1	STHA_SALTY
45	88	4.6	454	1	CBPH_LOPAM

ALIGNMENTS

RESULT 1
CBPB_RAT

ID CBPB_RAT STANDARD; PRT; 415 AA.

AC P19223;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Carboxypeptidase B precursor (EC 3.4.17.2).

GN CPB.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89034324; Pubmed=3182872;

RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;

RT "Structural characterization of the rat carboxypeptidase A1 and B

genes. Comparative analysis of the rat carboxypeptidase gene

family.";

RL J. Biol. Chem. 263:17837-17845(1988).

CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =

peptide + L-lysine (or L-arginine).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M23959; AAA40872.1;

DR EMBL; M23947; AAA40872.1; JOINED.

DR EMBL; M23950; AAA40872.1; JOINED.

DR EMBL; M23952; AAA40872.1; JOINED.

DR EMBL; M23953; AAA40872.1; JOINED.

DR EMBL; M23954; AAA40872.1; JOINED.

DR PIR; A32129; A32129.

DR HSPSP; P09955; INSA.

DR MEROPS; M14.003; -.


```
Db 259 AspSerSerProAenThrAsnLysProCysLeuAenValTyrArgGlyProAlaAlaProGlu 278
      ::::: ||| ||| |||
Qy 742 TCAGAACCAAGTGAAGCGAGTGGCTAGTTCTTTGAGAGAAATATCAACACGATTAA 801
      ||||| ||||| ::||| ::||| ::|||
Db 279 SerGluLysGluThrLysAlaValThrAsnPhelIleArgSerHisLeuAenSerIleLys 298
      ||||| ||||| ::||| ::||| ::|||
Qy 802 GCATACATCAGCATGCTATCATCCAGCATATAGTGTTCATATTCCTATACACGA 861
      ||||| ::||| ::||| ::||| ::|||
Db 299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318
      ||||| ::||| ::||| ::||| ::|||
Qy 862 AGTAAAGCAAGACCATGAGCACTGCTCTAGTAGCAGTGAAGCAGTTCGGTATT 921
      ||||| ::||| ::||| ::||| ::|||
Db 319 LysLeuProProAenHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338
      ||||| ::||| ::||| ::||| ::|||
Qy 922 GACAAAATAGTAAAAATACACAGGTATACATGCGCCATGCTCAGAAACCTCTATACCTA 981
      ||||| ||||| ::||| ::||| ::|||
Db 339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357
      ||||| ||||| ::||| ::||| ::|||
Qy 982 GCTCCTGGAGTGGGACGATGATCATGATTTGGGCATCAATATTCGTTTAC---- 1037
      ||||| ::||| ::||| ::||| ::|||
Db 358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377
      ||||| ::||| ::||| ::||| ::|||
Qy 1038 -----ATCAAAACCCACC 1049
      ||||| |||||
Db 378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397
      ||||| |||||
Qy 1050 TGTAGAGAGAGCTTTTGGCGCTGTC 1073
      ||||| |||||
Db 398 CysLysGluThrMetLeuSerVal 405
      ||||| |||||

RESULT 4
ID CBPC HUMAN STANDARD; PRT; 417 AA.
AC P15088; O96E94;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Lung;
RC SEQUENCE FROM N.A.
RX MEDLINE=90083291; PubMed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Mast cells;
RC MEDLINE=92105393; PubMed=1729276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A."
RL J. Clin. Invest. 89:273-282 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Bone marrow;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
```

```
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [4]
RP SEQUENCE OF 110-417 FROM N.A.
RX MEDLINE=92333165; PubMed=1629626;
RA Natsuaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuaki M.,
RA Wintroub B.U., Rutter W.J., Goldstein S.M.;
RT "Human skin mast cell carboxypeptidase: functional characterization,
RT cDNA cloning, and genealogy."
RL J. Invest. Dermatol. 99:138-145 (1992).
RN [5]
RP SEQUENCE OF 110-137.
RX MEDLINE=89214692; PubMed=2708524;
RA Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT "Human mast cell carboxypeptidase. Purification and
RT characterization."
RL J. Clin. Invest. 83:1630-1636 (1989)
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -!- SUBCELLULAR LOCATION: Secretory granules.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27717; AAA35652.1; -.
CC EMBL; M73720; AAA59568.1; -.
CC EMBL; M73716; AAA59568.1; JOINED.
CC EMBL; M73717; AAA59568.1; JOINED.
CC EMBL; M73718; AAA59568.1; JOINED.
CC EMBL; M73719; AAA59568.1; JOINED.
CC EMBL; BC012613; AAH12613.1; -.
CC EMBL; S40234; AAB22578.2; ALT_SEQ.
CC PIR; A43929; A43929.
CC HSPP; P09955; INSA.
CC Genew; HGNC:2298; CPA3.
CC MIM; 114851; -.
CC GO; GO:0030141; C:secretory granule; NAS.
CC GO; GO:0004182; F:proteolysis and peptidolysis; TAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
CC InterPro; IPR003146; Propep M14.
CC InterPro; IPR000834; Zn_carboPept.
CC Pfam; PF02244; Propep M14; 1.
CC Pfam; PF00246; Zn_carboPept; 1.
CC SMART; SM00631; Zn_pept; 1.
CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
CC Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
KW SIGNAL
FT PROPEP 1 15 ACTIVATION PEPTIDE.
FT CHAIN 16 109 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 110 417 ZINC (BY SIMILARITY).
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
```

FT DISULFID 173 186 BY SIMILARITY. 417
FT DISULFID 245 268 BY SIMILARITY. 417
FT CONFLICT 63 N -> K (IN REF. 31). 139
SQ SEQUENCE 417 AA; 48700 MW; 8CB90DB758117B24 CRC64; 132

Alignment Scores:
Pred. No.: 7.54e-49 Length: 417
Score: 635.00 Matches: 139
Percent Similarity: 52.31% Conservative: 76
Best Local Similarity: 33.82% Mismatches: 132
Query Match: 32.87% Indels: 65
DB: 1 Gaps: 9

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPC_HUMAN (1-417)

QY 16 CTTGAGTCCTTGTACCATTTGCTCTTCTGTGAGCAGCATGCTTCGCG----- 66
DB 1 MetArgLeuIleLeuProValGlyLeuIleAlaThrLeuAlaIleAlaProValArg 20
QY 67 TTCAGAGTGGCCCAAGTTCTAGCTCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
DB 21 PheAspArgGluLeuValPheArgValysProGlnAspGluLeuGlnAlaAspIle 40
QY 127 CAGATCTTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGACCTTATT 186
DB 41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisVal 60
QY 187 GTGAAGAAAAAACAAGTCCATTTTGTAAATGCTATGTCGCAATGTGAAGCC 246
DB 61 AlaAlaAsnMetMetValAspPheArgValSerGluLeuGlnAlaIleGlnSer 80
QY 247 CATTTAAATGTGAGGAATTCATGCTGCTGCTGCGCAGCGGTAACAGCTGACCTTATT 306
DB 81 AlaLeuAspGlnAsnLysMetHisTyrGluLeuLeuIleHisAspLeuGlnGluLeu 100
QY 307 CAACGACGAGATTTC---AAGCACACAGTCAGCCCGAGCTCCGCGATCGTACTATGAA 363
DB 101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer-----TyrAla 117
QY 364 CAGTATCTCTCACTAAATGAATCTATCTTGGATAGATTTATTAACAGAGCATCTCT 423
DB 118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137
QY 424 GATATGCTTCAAAAATCCATTTGATCCCTCAITTCAGAGAGTACCCACTCTATGTTTAA 483
DB 138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157
QY 484 AAGCTTTTCTGAAAGAACAAACAGCCCAAAATCCATATGATTTGATGCTGGGAATCCAT 543
DB 158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176
QY 544 GCCAGAGATGATCTCTCTGCTTTCTGCTGTTGTTGTTTCATA----- 585
DB 177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196
QY 585 ----- 585
DB 197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216
QY 586 -----GGCCAT-----AATCGAATGTGGAGAAAGAAC 612
DB 217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236
QY 613 CGTTCTTCTATGCGAACATCATTTGATCGGACAGACCTGAATAGCACTTTGTCTCC 672
DB 237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255
QY 673 AAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTGCTCGGAACCTACTGCTGGACTT 732
DB 256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgLysSer 275
QY 733 TATCCTGAGTCAGAACCAAGAGTGAAGGCGAGTGGCTAGTTCTTCTGAGAGAAATATCAAC 792
DB 733 TATCCTGAGTCAGAACCAAGAGTGAAGGCGAGTGGCTAGTTCTTCTGAGAGAAATATCAAC 792

276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295
793 CAGATAAAGCATACATCAGCATGCTATCTATCTCCAGCATATAGTGTTCATATTC 852
296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315
853 TATACACGAAAGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
316 TyrThrSerLysLeuProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335
913 CGTGTCTATTGACAAACTAGTAAATACACAGGTATACATGCGCCTGCTGGAAC 972
336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354
973 TTATACCTAGCTCTCTGAGGTGGGAGGATTTGATCTATGATTTGGSCATCAATATTCG 1032
355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374
1033 TTTTAC-----ATC 1040
375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394
1041 AAACCCACTGTAGAGAGCTTTTGGCGTGTCT 1073
395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 5
CBPB_PIG STANDARD; PRT; 401 AA.
AC P09955;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CBP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE OF 1-103.
RX MEDLINE=91208150; PubMed=2018774;
RA Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA Aviles F.X.;
RT "Analysis of the activation process of porcine procarboxypeptidase B
and determination of the sequence of its activation segment.";
RL Biochemistry 30:4082-4089 (1991).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-38.
RX MEDLINE=85279427; PubMed=4026847;
RA Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.,
RT "Sequential homologies between procarboxypeptidases A and B from
porcine pancreas.";
RL Biochem. Biophys. Res. Commun. 130:97-103 (1985).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=91114690; PubMed=1989878;
RA Coll M., Guasch A., Aviles F.X., Huber R.;
RT "Three-dimensional structure of porcine procarboxypeptidase B: a
structural basis of its inactivity.";
RL EMBO J. 10:1-9 (1991).
RN [4]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX MEDLINE=91027767; PubMed=2223783;
RA Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RT "Sequence-specific 1H NMR assignments and determination of the
secondary structure for the activation domain isolated from
pancreatic procarboxypeptidase B.";
RL Biochemistry 29:7515-7522 (1990).
RN [5]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX MEDLINE=91114693; PubMed=1989879;
RA Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;

"The NMR structure of the activation domain isolated from porcine procarboxypeptidase B."; EMBO J. 10:11-15(1991). [6]

STRUCTURE BY NMR OF ACTIVATION PEPTIDE.

RA MEDLINE=93044373; PubMed=1422143;
RA Billeter M., Vendrell J., Wider G., Aviles P.X., Coll M., Guasch A.,
RA Huber R., Wuthrich K.;
RT "Comparison of the NMR solution structure with the X-ray crystal
RT structure of the activation domain from procarboxypeptidase B.";
RL J. Biomol. NMR 2:1-10(1992).
CC J.-I. CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -I- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COB.html".

PDB; 1PBA; 31-OCT-93.

PDB; 1NSA; 24-DEC-97.

MEROPS; M14.003; -

InterPro; IPR003146; Propep M14.

InterPro; IPR000834; Zn carboxypept.

Pfam; PF02244; Propep M14; 1.

Pfam; PF00246; Zn carboxypept; 1.

PRINTS; PR00765; CRBOXPTASEA.

SMART; SM00631; Zn_pept; 1.

PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.

PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.

KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;

KW 3D-structure.

PROPEP 1 95 ACTIVATION PEPTIDE.

CHAIN 96 401 CARBOXYPEPTIDASE B.

FT DISULFID 158 171

FT DISULFID 230 253

FT DISULFID 244 258

FT METAL 161 161

FT METAL 164 164

FT METAL 289 289

FT ACT SITE 341 341

FT ACT SITE 363 363

FT TURN 8 9

FT STRAND 11 17

FT HELIX 20 32

FT STRAND 36 39

FT HELIX 43 45

FT STRAND 50 56

FT HELIX 58 60

FT STRAND 61 69

FT TURN 70 72

FT STRAND 74 79

FT HELIX 82 87

FT TURN 88 88

FT TURN 101 102

FT HELIX 107 120

FT TURN 122 124

FT STRAND 125 132

FT TURN 134 135

FT STRAND 138 144

FT TURN 149 150

FT STRAND 153 157

FT TURN 162 163

FT HELIX 165 181

FT TURN 182 184

FT HELIX 186 194

FT STRAND 196 200

FT HELIX 205 213

FT TURN 214 214

FT STRAND 216 217

FT TURN 224 224

FT STRAND 226 227

FT TURN 231 231

FT STRAND 235 237

FT TURN 243 244

FT STRAND 249 249

FT TURN 253 253
FT TURN 256 256
FT STRAND 257 257
FT TURN 262 263
FT HELIX 266 277
FT TURN 278 281
FT STRAND 282 289
FT TURN 294 297
FT STRAND 307 308
FT HELIX 309 327
FT STRAND 332 335
FT HELIX 336 339
FT HELIX 347 353
FT TURN 354 355
FT STRAND 359 363
FT TURN 371 372
FT HELIX 376 378
FT HELIX 379 398
FT TURN 399 401
SQ SEQUENCE 401 AA; 53129AF159A26348 CRC64;

Alignment Scores:

Pred. No.: 2.32e-48 Length: 401
Score: 629.50 Matches: 134
Percent Similarity: 54.85% Conservative: 64
Best Local Similarity: 37.12% Mismatches: 120
Query Match: 32.58% Indels: 43
DB: 1 Gaps: 6

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPB_PIG (1-401)

Qy 67 TTCAGAGTGCACCAAGTTCTAGCTCTTCTAGAACCTCTTAGCGAAGTTCAGTTCTA 126
Db 7 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerGluLeu 26
Qy 127 CAGAACTTTACTCAACATATGAGATTGTTCTTCGCAGCGCGGTAAACAGCTGACCTATT 186
Db 27 HisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGlnIle 46
Qy 187 GTGAAGAAAAACAAGTCCATTTTGTGTAATGCATCTCATGTCGACAAATGTGAAAGCC 246
Db 47 LysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGluAsp 66
Qy 247 CATTTAAATGTGAGCGGAATTCATGTCAGTGTCTTCGTCGACAGCGTGAAGATCTTTATT 306
Db 67 PheLeuGluGlnAsnGluLeuGlnTyrGluValIleAsnLeuAsnLeuArgSerValLeu 86
Qy 307 CAACAGCAGATTTCCAAACGACACAGTCAGCCCGCGCCCTCCGATCGTACTATGAACAG 366
Db 87 GluAlaGlnPheAspSerArgCys-----ArgThrThrGlyHisSerTyrGluLys 103
Qy 367 TATCACTCACTAAATGAATCTATTCTTGATAGATTATTAAGTACGAGAGCGCATCTTGAT 426
Db 104 TyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValThrSerLysAsnProAsp 123
Qy 427 ATGCTTACAAAAATCCACATTCGATTCCTATTTGAGAAGTACCCACTCTCTATGTTTAAAG 486
Db 124 LeuIleSerArgSerAlaIleGlyThrThrPheAspGlyAspAsnIleTyrLeuLeuLys 143
Qy 487 GTTCTCGAAAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
Db 144 Val---GlyLysProGlySerAsnLysProAlaIlePheMetAspCysGlyPheHisAla 162
Qy 547 AGAGAATGGAATCTCTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 585
Db 163 ArgGluTrpIleSerGlnAlaPheCysGlnTrpPheValArgAspAlaValArgThrTyr 182
Qy 585 ----- 585
Db 183 GlyTyrGluAlaHisMetThrGluPheLeuAspAsnLeuAspPheTyrValLeuProVal 202
Qy 586 -----GGCCAT-----AATCGAATGTGGAGAAAGACCGT 615
|||||

Db 203 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThrArg 222
Qy 616 TCCTTCTATGCGAACATCATTGTCATGCGACACACCTGATACCACTTGTCTCCAAA 675
Db 223 SerThrAsnAlaGlySerSerCysThrGlyThrAspProAsnArgAsnPhe---AsnAla 241
Qy 676 CACTGGTGTGAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
Db 242 GlyTrpCysThrValGlyAlaSerValAsnProCysAsnGluThrTyrCysGlySerAla 261
Qy 736 CCGTGAAGCAGAACAGAGTGAAGCGAGTGTCTTCTTGTGAGAGAAATATCAACAG 795
Db 262 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSerSer 281
Qy 796 ATTAAGCATACATCAGCATCATCTATCTACCTCCAGCATATAGTGTTCCTATTCCTAT 855
Db 282 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 301
Qy 856 ACACGAAGTAAAGCAAGACCATGAGCACTGTCTCTAGTACGACGTGAAGCAGTTCGT 915
Db 302 AspTyrLysLeuProGluAsnAspAlaGluLeuAsnSerLeuAlaLysGlyAlaValLys 321
Qy 916 GCTATTGACAAACTAGTAAATAATACAGGATATACATGCGCCATGCTCAGAACTTA 975
Db 322 GluLeu---AlaSerLeuTyrGlyThrSerTyrSerTyrGlyProGlySerThrThrIle 340
Qy 976 TACCTAGCTCTGAGGCGGAGTGGATCTCATGATTTGGGCATCAATATTCGTTT 1035
Db 341 TyrProAlaAlaGlySerAspTrpAlaTyrAsnGlnGlyIleLysTyrSerPhe 360
Qy 1036 ACA 1038
Db 361 Thr 361

RESULT 6
ID CBPB CANFA STANDARD; PRT; 416 AA.
AC P55261,
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
GN membrane associated protein) (ZAP47).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Fukuoka S.-I.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2O) =
CC peptide + L-lysine (or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D78348; BRA11366.1; -.
DR HSSP; P09955; 1PBA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn carboxypept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn carboxypept; 1.
DR PRINTS; PR00765; CRBOXPTASEA.

DR SMART; SM00631; Zn pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109
FT CHAIN 110 416
FT METAL 175 175
FT METAL 178 178
FT METAL 303 303
FT ACT SITE 377 377
FT DISULFID 244 267
FT DISULFID 258 272
SQ SEQUENCE 416 AA; 47595 MW; DFE1992CD52F8AB4 CRC64;
Alignment Scores:
Pred. No.: 3,52e-48 Length: 416
Score: 627.50 Matches: 140
Percent Similarity: 51.10% Conservative: 69
Best Local Similarity: 34.23% Mismatches: 133
Query Match: 32.48% Indels: 68
DB: 1 Gaps: 9
US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPB_CANFA (1-416)
Qy 22 GTCTTGTACCCATTGTCTC-----TTCTGTGAGCAGCATGTCTTCGGGTTTC 69
Db 5 IleLeuValThrLeuAlaLeuAlaSerAlaHisTyrSerGlyGluHis-----Phe 21
Qy 70 CAGATGGCCCAAGTTCTAGCTGTCTTCTCTAGAACCTCTTAGGCAAGTCAAGTTCTACAG 129
Db 22 GluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnLeuLeuHis 41
Qy 130 AATCTTACTACAAATATGATGTTCTCTGCGACCGGTAAACAGCTGACCTTATTGTG 189
Db 42 ThrLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLys 61
Qy 190 AAGAAAAACAAGTCCATTTTGTAAATGCATCTGTCGACATGTGCAATGTGAAGCCCAT 249
Db 62 ProHisSerThrAlaAspPheArgValLysAlaGluAspIleLeuThrValGluAspPhe 81
Qy 250 TTAATGTGAGCGGAATTCATCGACGTCTCTGCGACAGCGTGAAGATCTTATTCAA 309
Db 82 LeuLysGlnAsnGluLeuHisTyrGluValLeuIleAsnAsnLeuArgLeuValLeuGlu 101
Qy 310 CAGCAGATTTCCAAACGACACAGTCAGCCCGGAGCTCCCGCATCTACTATGAACAGTAT 369
Db 102 GlyGlnPheGlyArgGlnVal-----ProAlaThrGlyHisSerTyrGluLysTyr 118
Qy 370 CACTCACTAAATGAATCTATTCTTGTGATAGATTTTATACTGAGGACATCTGTATG 429
Db 119 AsnArgTrpGluThrIleGluAlaTrpThrGlnGlnValThrSerGluAsnProAspLeu 138
Qy 430 CTTACAAAAATCCACATTTGATGCTCTCATTTGAGAAGTACCACCTATGTTTAAAGTT 489
Db 139 IleSerArgArgSerIleGlyThrThrPheGluGlyArgThrIleTyrLeuLeuLeuVal 158
Qy 490 TCTGGAAGAACAAACAGCCAAAATGCCATATGATGATTCATGATTCATCCATGCCAGA 549
Db 159 ---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArg 177
Qy 550 GAATGATCTCTCTGCTTCTCTGCTGCTGCTCATA----- 585
Db 178 GluTrpIleSerProAlaPheTrpGlnTrpPheValArgGlu***IleArgThrTyrGly 197
Qy 585 ----- 585
Db 198 GlnGluIleHisMetThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValGly 217
Qy 586 -----GGCCAT-----AATCGAATGTGGAGAAAGACCGTTCT 618
Db 218 AsnIleAspGlyTyrValTyrThrTrpThrLysAsnArgMetTrpArgLysThrArgSer 237


```

Db 145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163
QY 727 GGACTTTATCTGAGTCAGAACCAAGTGAAGCAGTCGGCTAGCTTTCTTTCAGAGAAT 786
Db 164 GlySerAlaAlaGluSerGluGlySerLysAlaValAlaAspPheIleArgAsnHis 183
QY 787 ATCAACAGATTAAAGATACATCAGCATGATTCATACACTCCAGCATATAGTGTTCCTCA 846
Db 184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrPro 203
QY 847 TATTCCTATACACGAAGTAAAGCAACACATCAGGAAGTGTCTCTAGTAGCCAGTAA 906
Db 204 TyrSerTyrAspTyrLysLeuProLysAsnValGluLeuAsnThrLeuAlaLysGly 223
QY 907 GCAGTTCGTCTATTGACAAACTAGTAAATACCAAGTATACATACATGCGCTGCTCA 966
Db 224 AlaValLysLysLeu---AlaSerLeuHisGlyThrTyrSerTyrGlyProGlyAla 242
QY 967 GAAACCTTATACCTAGTCTCTGGAGGTGGGACGATGGATCTATGATTTGGGCATCAAA 1026
Db 243 ThrThrIleTyrProAlaSerGlyGlySerAspSerTrpAlaTyrAspGlnGlyLys 262
QY 1027 TATTCGTTTATCATC----- 1040
Db 263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282
QY 1041 -----AAACCCACTGTAGAGAGCTTTGCCGCTGTC 1073
Db 283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIle 295

RESULT 8
CBPC RAT STANDARD; PRT; 309 AA.
ID _CBPC RAT
AC P21961;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA Neurath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules.";
RL Biochemistry 30:648-655(1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A38395; A38395.
DR HSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR InterPro; IPR000834; Zn carboxopept.
DR Pfam; PF00246; Zn carboxopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
KW METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 196 196 ZINC (BY SIMILARITY).
FT ACT SITE 248 248 PROTON DONOR (BY SIMILARITY).
FT ACT SITE 270 270 NUCLEOPHORE (BY SIMILARITY).
FT DISULFID 65 78 BY SIMILARITY.
FT DISULFID 137 160 BY SIMILARITY.

```

```

FT VARIANT 1 1 MISSING (IN THE MAJOR FORM).
SQ SEQUENCE 309 AA; 35786 MW; 20330FABC3EE83EF CRC64;
Alignment Scores:
Pred. No.: 9.03e-42 Length: 309
Score: 555.50 Matches: 118
Percent Similarity: 54.27% Conservative: 41
Best Local Similarity: 40.27% Mismatches: 77
Query Match: 28.75% Indels: 58
DB: 1 Gaps: 6
US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPC_RAT (1-309)
QY 358 TATGACAGTATCACTCACTAAATGAATCTATTCTTTGGATAGAAATTTATACTGAGAGG 417
Db 8 TyAlaLysTyrAsnAspTrpAsnLysIleValSerTyrThrGluLysMetValGluLys 27
QY 418 CATCTGATATGCTTCAAAAATCCACATTCGATTCCTCATTTGAGAAGTACCCACTCTAT 477
Db 28 HisProGluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyr 47
QY 478 GTTTTAAAGTTTCTGAAAAGAAACAAACGACCAAAATGCCATATGGATGTGATGGA 537
Db 48 ValLeuLysIle---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66
QY 538 ATCCATGCGCAGAGAATGGATCTCTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 585
Db 67 IleHisAlaArgGluIrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86
QY 586 -----GGCCATAAT----- 594
Db 87 LysSerTyrGlyLysAsnAsnIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106
QY 595 -----CGAATGTGAGA 606
Db 107 LeuProValPheAsnValAspGlyTyrIleTrpSerTyrThrLysAspArgMetTrpArg 126
QY 607 AAGAACCGTTTCTTCTATGCGAACATCATTCGATCGAAGAGGAGTGGCTAGTCTTCTT 666
Db 127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146
QY 667 GTCTCAAAACACATGCTGTGAGGAAGGTGCATCCAGTTCCTCATCTCGGAAACCTACTGT 726
Db 147 ---AspValSerTrpAspSerSerProAsnThrAspAsnProCysLeuSerValTyrArg 165
QY 727 GGACTTTATCTGAGTCAGAACCAAGTGAAGCAGTCGGCTAGCTTTCTTTCAGAGAAT 786
Db 166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185
QY 787 ATCAACAGATTAAAGATACATCAGCATGATTCATACACTCCAGCATATAGTGTTCCTCA 846
Db 186 LeuAsnSerIleLysAlaTyrLeuThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205
QY 847 TATTCCTATACACGAAGTAAAGCAACACATCAGGAAGTGTCTCTAGTAGCCAGTAA 906
Db 206 TyrGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLysValAlaArgIle 225
QY 907 GCAGTTCGTCTATTGACAAACTAGTAAATACCAAGTATACATGCGCTGCTGCTCA 966
Db 226 AlaThrAspValLeu---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAla 244
QY 967 GAAACCTTATACCTAGTCTCTGAGGTGGGACGATGGATCTATGATTTGGGCATCAAA 1026
Db 245 SerThrIleTyrLysThrSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyLys 264
QY 1027 TATTCGTTTATC----- 1037
Db 265 HisThrPheAlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSer 284
QY 1038 -----ATCAAAACCCACTGTAGAGAAGCTTTTGGCGCTGTC 1073
Db 285 ArgIleLysProThrCysLysGluThrMetLeuSerVal 297

```

RESULT 9
 CBP2 RAT
 ID CBP2 RAT STANDARD; PRT; 417 AA.
 AC P19222;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
 GN CPA2.
 OS Rattus norvegicus (Rat).
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RN TISSUE=Pancreas;
 RN MEDLINE=89034323; PubMed=3182871;
 RA Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
 RA Graf W., Rutter W.J.;
 RT "A novel rat carboxypeptidase, CPA2: characterization, molecular
 RT cloning, and evolutionary implications on substrate specificity in
 RT the carboxypeptidase gene family.";
 RL J. Biol. Chem. 263:17828-17836(1988).
 RN [2]
 RN SEQUENCE OF 131-143 FROM N.A.
 RN MEDLINE=95386501; PubMed=7657630;
 RA Normant E., Gros C., Schwartz J.C.;
 RT "Carboxypeptidase A isoforms produced by distinct genes or
 RT alternative splicing in brain and other extrapancreatic tissues.";
 RL J. Biol. Chem. 270:20543-20549(1995).
 RN [3]
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RN MEDLINE=92105124; PubMed=1761558;
 RA Faming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
 RT "Structural evolution of an enzyme specificity. The structure of rat
 RT carboxypeptidase A2 at 1.9-A resolution.";
 RL J. Biol. Chem. 266:24606-24612(1991).
 CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
 CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
 CC residues.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M23721; AAA40956.1; -
 CC EMBL; M23714; AAA40956.1; JOINED.
 CC EMBL; M23715; AAA40956.1; JOINED.
 CC EMBL; M23716; AAA40956.1; JOINED.
 CC EMBL; M23717; AAA40956.1; JOINED.
 CC EMBL; M23718; AAA40956.1; JOINED.
 CC EMBL; M23719; AAA40956.1; JOINED.
 CC EMBL; M23720; AAA40956.1; JOINED.
 CC EMBL; S79837; -; NOT_ANNOTATED_CDS.
 CC PIR; A32128; A32128.
 CC HSSP; P48052; IAYE.
 CC InterPro; IPR003146; Propep M14.
 CC InterPro; IPR000834; Zn_carboxypeptidase.
 CC Pfam; PF02244; Propep M14; 1.
 CC Pfam; PF00246; Zn_carboxypeptidase; 1.
 CC PRINTS; PR00765; CRBOXYPTASEA.
 CC SMART; SM00631; Zn_pept; 1.
 CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 CC PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 16 POTENTIAL.
 FT PROPEP 17 112 ACTIVATION PEPTIDE.
 FT

Db 309 SerGlnLeuLeuLeuTyProTyGlyThrThrGlnSerIleProAspLysThrGlu 328
 Qy 886 CTGCTCTAGTAGCAGTGAAGCAGTTCGCTATTGACAAACACTAGTAAATAATACCAGG 945
 Db 329 LeuGlnValAlaLysSerAlaValGluAlaLeu---LysSerLeuTyGlyThrSer 347
 Qy 946 TATACATGATGCCATGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGGAGCATGG 1005
 Db 348 TyrlsTyGlySerIleIleThrThrIleTyGlnAlaSerGlyGlySerIleAspTip 367
 Qy 1006 ATCTATGATTTGGGCATCAATATTCGTTTACA 1038
 Db 368 SerTyrlsGlnGlyIleLysTySerPheThr 378
 RESULT 11
 CBP1_RAT STANDARD; PRT; 419 AA.
 ID CBP1_RAT
 AC P00731;
 DT 21-JUL-1996 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPA1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82105986; PubMed=6275388;
 RA Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
 RA Pictet R.L., Valenzuela P., Rutter W.J.;
 RA "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
 RT characterization of the gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89034324; PubMed=3182872;
 RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
 RA "Structural characterization of the rat carboxypeptidase A1 and B
 RT genes. Comparative analysis of the rat carboxypeptidase gene
 RT family.";
 RL J. Biol. Chem. 263:17837-17845(1988).
 CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
 CC amino acid.
 CC -1- SUBUNIT: Monomer.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; V01232; CAA24542.1; -;
 DR EMBL; J00713; AAA40893.1; -;
 DR EMBL; M23990; AAA40955.1; -;
 DR EMBL; M23960; AAA40955.1; JOINED.
 DR EMBL; M23985; AAA40955.1; JOINED.
 DR EMBL; M23986; AAA40955.1; JOINED.
 DR EMBL; M23987; AAA40955.1; JOINED.
 DR EMBL; M23988; AAA40955.1; JOINED.
 DR EMBL; M23989; AAA40955.1; JOINED.
 DR PIR; A00911; CPRTA.
 DR HGSP; P00730; 1PVT.
 DR MEROPS; M14.001; -;
 DR InterPro; IPR003146; Propep M14.
 DR InterPro; IPR000834; Zn_carbopept.
 DR Pfam; PF02244; Propep M14; 1.
 DR Pfam; PF00246; Zn_carbopept; 1.
 DR SMART; SM00631; Zn_pept; 1.

DR PROSITE; PS00132; CARBOXYPEPT ZN 1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT ZN 2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 110 ACTIVATION PEPTIDE.
 FT PROPEP 111 419 CARBOXYPEPTIDASE A1.
 FT METAL 179 179 ZINC (BY SIMILARITY).
 FT METAL 182 182 ZINC (BY SIMILARITY).
 FT METAL 306 306 ZINC (BY SIMILARITY).
 FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 248 271 BY SIMILARITY.
 FT CONFLICT 196 196 I -> V (IN REF. 1).
 FT CONFLICT 261 263 FGM -> LGK (IN REF. 1).
 FT CONFLICT 347 347 K -> E (IN REF. 1).
 SQ SEQUENCE 419 AA; 47197 MW; BB002D1CB9987491 CRC64;
 Alignment Scores:
 Pred. No.: 1.99e-39 Length: 419
 Score: 529.50 Matches: 134
 Percent Similarity: 50.77% Conservative: 63
 Best Local Similarity: 34.54% Mismatches: 136
 Query Match: 27.41% Indels: 55
 DB: 1 Gaps: 10
 US-09-980-881A-1_COPY_18_1097 (1-1080) x CBP1_RAT (1-419)
 Qy 4 AGCTTTGACGCTTCGAGTCTTGTACCCATTGTTCTCTTGTGTGAGCAGCATGTCTTC 63
 Db 3 ArgLeuLeuIleLeuSerLeuLeuGluAlaValCysGlyAsnGluAen----- 19
 Qy 64 GCCTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTTAGAACCTCTAGGCAAGTCAAGTT 123
 Db 20 ---PheValGlyHisGlnValLeuArgIleSerAlaAlaAspGluAlaGlnValGlnLys 38
 Qy 124 CTACAGAACTTCTACTACACATATGAGATTGTCTC-----TGGCAGCCGCGTAACAGCT 177
 Db 39 ValLysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgAspAlaAlaArg 58
 Qy 178 GACCTTATGTGAGAGAAAACCAAGTCCATTTTGTAAATGATCATGTGTCGACAAT 237
 Db 59 AlaGlyIleProIleAspValArgValProPhe-----ProSerIleGlnSer 74
 Qy 238 GTGAAGACCCATTTAAATGTGAGCGAATTCATGAGTGTCTGTCGCGAGAGTGGAA 297
 Db 75 ValLysAlaPheLeuGluTyHisGlyIleSerTyTrpGluIleMetIleGluAspValGln 94
 Qy 298 GATCTTTATT-----CAACAGCAGATTTCACACGACACAGCATGACCCCGAGCTCC 348
 Db 95 LeuLeuLeuAspGluGluLysGlnMetSerAlaPheGlnAlaArgAlaLeuSerThr 114
 Qy 349 GCATCGTAC---TATGACAGTATCATCTAAATGAATCTTCTTGTGATGAGATTT 405
 Db 115 AspSerPheAsnTyAlaThrTyHisThrLeuAspGluIleTyTrpGluPheMetAspLeu 134
 Qy 406 ATAACAGAGGATCTGATATGCTTACAAAATCCACATTCGATCTCTATTGACAAG 465
 Db 135 LeuValAlaGluHisProGlnLeuValSerLysIleGlnIleGlyAsnThrPheGluGly 154
 Qy 466 TACCCACTATGTTTAAAGGTTTCTGAAAAGACAAACAGCCCAAAATGCCATATGG 525
 Db 155 ArgProIleHisValLeuLysPheSer---ThrGlyGlyThrAsnArgProAlaIleTip 173
 Qy 526 ATTGACTGTGGAATCCATGCGAGAGATGATCTCTCTGCTTCTGCTGTTGTTGTTTCA 585
 Db 174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193
 Qy 586 GGCCAT----- 591
 Db 194 LysLysIleThrLysAspTyTrpGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213
 Qy 592 -----AAT 594
 |||

Alignment Scores:

Pred. No.: 8.91e-38 Length: 417
 Score: 511.00 Matches: 128
 Percent Similarity: 48.73% Conservative: 64
 Best Local Similarity: 32.49% Mismatches: 136
 Query Match: 26.45% Indels: 66
 DB: 1 Gaps: 11

US-09-980-881A-1_COPY_18_1097 (1-1080) x CPB2_HUMAN (1-417)

Qy 55 CATGCTTC-----GCCTTCAGAGTGGCCAGTTCAGTCTTCCTAGAAC 105
 |||:|||||
 Db 13 HisiletyrCysLeuGluThrPheValGlyAspGlnValLeuGluLeuValProSerAsn 32
 |||:|||||
 Qy 106 TCTAGGCAAGTTCAGTCTACAGAACTTACTACACATATGAGATGTTCTC----- 159
 |||:|||||
 Db 33 GluGluGlnIleuYAsnLeuLeuGlnLeuGluAlaGlnGluHisLeuLeuAspPhe 52
 |||:|||||
 Qy 160 TGGCAGCGGTAAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAAT 219
 |||:|||||
 Db 53 TrpLysSerProThr-----ThrProGlyGluThrAlaHisValArgValPro 68
 |||:|||||
 Qy 220 GCATCTGATGCGCAATGTGAAGCCCATTAATGTGAGCGGAATTCACGAGTGT 279
 |||:|||||
 Db 69 PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaIleSerIle 88
 |||:|||||
 Qy 280 TTGCTGGCAGCGTGAAGATCTTATTCAACAGCAGATTTCCACGACACAGTCAGCCCC 339
 |||:|||||
 Db 89 MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg 108
 |||:|||||
 Qy 340 CGAGCCTCCGATCG-----TACTATGAACAGATCACTCACTAAATGAATCTAT 390
 |||:|||||
 Db 109 ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaIleThrLeuGluGluLeuSer 128
 |||:|||||
 Qy 391 TCTTGGATAGAAATTTATACTGAGAGGATCCTCATATGCTTACAAAATCCACATGGA 450
 |||:|||||
 Db 129 GlnGluMetAspAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly 148
 |||:|||||
 Qy 451 TCCTCATTTGAGAGTACCCTACTATGTTTAAAGGTTCT---GGAAAGACAACA 507
 |||:|||||
 Db 149 SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyAsp----- 166
 |||:|||||
 Qy 508 GCCAAAATGCATATGATGATGCTGGAAATCCATCCAGAGATGGATTCCTCTGCT 567
 |||:|||||
 Db 167 ---LysProAlaIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla 185
 |||:|||||
 Qy 568 TTCTGCTGTGGTTCATAGGC----- 588
 |||:|||||
 Db 186 ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205
 |||:|||||
 Qy 588 ----- 588
 |||:|||||
 Db 206 SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal 225
 |||:|||||
 Qy 589 -----CATAATCGAATGTGGAAGAAGAACCGTTCTTCTATGCGAACCAATCAT 636
 |||:|||||
 Db 226 PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLysValSerGlySerLeu 245
 |||:|||||
 Qy 637 TGCATCGGAACAGACCTGAATAGCAACTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 696
 |||:|||||
 Db 246 CysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla 264
 |||:|||||
 Qy 697 TCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCTGTAGTCAGACACGAGGTG 756
 |||:|||||
 Db 265 SerSerAsnProCysSerAspSerTyrHisGlyProSerAlaAsnSerGluValGluVal 284
 |||:|||||
 Qy 757 AAGCAGTGGCTAGTTCCTCAGAGAAGAAATCAACAGATTAAAGCATATCAGCATG 816
 |||:|||||
 Db 285 LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleThrIle 303
 |||:|||||
 Qy 817 CATTCATCTCCAGCATATAGTGTTCATPATTTCTTATACACGAGTAAAGCAAGAC 876
 |||:|||||
 Db 304 HisSerTyrSerGlnLeuMetPheProTyrGlyTyrLysCysThrLysLeuAspAsp 323
 |||:|||||

Qy 877 CATGAGGAAGTCTCTAGTAGCAGTGAAGCAGTTCGTCTATTGACAAAACCTAGTAAA 936
 |||:|||||
 Db 324 PheAspGluLeuSerGluValAlaGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis 342
 |||:|||||
 Qy 937 AATACACAGGTATACACATGCGCCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGG 996
 |||:|||||
 Db 343 GlyThrLysTyrLysValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySer 362
 |||:|||||
 Qy 997 GACGATGGATCTATGATTTGGGCATCAATATATCGTTTACATCAAC----- 1044
 |||:|||||
 Db 363 IleAspTrpSerTyrAspTyrGlyIleLysTyrSerPheAlaPheGluLeuArgAspThr 382
 |||:|||||
 Qy 1045 -----CCACCTGTAGAGAAGCTTTTGGCG 1068
 |||:|||||
 Db 383 GlyArgTyrGlyPheLeuLeuProAlaArgGlnIleLeuPro 396
 |||:|||||

RESULT 13
 CBP1_HUMAN STANDARD; PRT; 419 AA.
 AC P15085; Q9BS67;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DB Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPA1 OR CPA
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=93038569; PubMed=1417781;
 RA Catusas L., Villegas V., Pascual R., Aviles F.X.,
 RA Wicker-Planquart C., Puigserver A.;
 RT "cDNA cloning and sequence analysis of human pancreatic
 RT procarboxypeptidase A1".
 RL Biochem. J. 287:299-303(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=22398257; PubMed=12477932;
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Atschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 17-42.
 RX MEDLINE=89153096; PubMed=2920728;
 RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
 RT "Purification and properties of five different forms of human
 RT procarboxypeptidases".
 RL Eur. J. Biochem. 179:609-616(1989).
 CC -I- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
 CC amino acid.
 CC -I- SUBUNIT: Monomer.

CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X67318; CAA47732.1; -.

DR EMBL; BC005279; AA05279.1; -.

DR PUR; S29127; S29127.

DR HSP; P00730; LPYT.

DR MEROPS; M14.001; -.

DR Genew; HGNC:2296; CPA1.

DR MIM; 114850; -.

DR GO; GO:0004182; F:carboxypeptidase A activity; TAS.

DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.

DR InterPro; IPR003146; Proprep M14.

DR Pfam; PF02244; Proprep M14; 1.

DR Pfam; PF02246; Zn carboxypeptidase.

DR PRINTS; PR00765; CRBOXYPTASEA.

DR SMART; SM00631; Zn_pept; 1.

DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.

DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.

DR Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.

FW SIGNAL 1 16

FT PROPEP 17 110 ACTIVATION PEPTIDE.

FT CHAIN 111 419 CARBOXYPEPTIDASE A1.

FT METAL 179 179 ZINC (BY SIMILARITY).

FT METAL 182 182 ZINC (BY SIMILARITY).

FT METAL 306 306 ZINC (BY SIMILARITY).

FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).

FT DISULFID 248 271 BY SIMILARITY.

FT CONFLICT 276 276 H -> R (IN REF. 2).

SQ SEQUENCE 419 AA; 47140 MW; 439FAFFFAEE958B1 CRC64;

Alignment Scores:

Pred. No.: 7.42e-36 Length: 419

Score: 489.50 Matches: 131

Percent Similarity: 47.63% Conservative: 60

Best Local Similarity: 32.67% Mismatches: 127

Query Match: 25.34% Indels: 83

DB: 1 Gaps: 13

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBP1_HUMAN (1-419)

Qy 7 CTTTGAGCCTTGAGTCTGAGTCCATGTTCTCTCTGAGCAGCATGTTCTCGCG 66

Db 4 LeuLeuValLeuSerValLeuLeuGlyAlaValPheGlyLysGluasp----- 19

Qy 67 TTCAGAGTGGCAGTCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAGTTCTA 126

Db 20 PheValGlyHisGlnValLeuArgIleSerValAlaAspGluAlaGlnValGlnLysVal 39

Qy 127 CAGNATCTTACTACAAATATGATGTTCTC-----TGGCAG----- 165

Db 40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProAlaHisPro 59

Qy 166 -----CCGTAACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTGTGTAAT 219

Db 60 GlySerProIleaspval-----ArgValProPhe----- 69

Qy 220 GCATCTGATGCAATGTGAAGCCCATTTAAATGTAGCGGAATTCATGCGAGTGC 279

Db 70 ---ProSerIleGlnAlaValLysIlePheLeuGluSerHisGlyIleSerGlyThr 88

Qy 280 TTGCTGCGACAGCTGGAGATCTTATT-----CAACAGCAGATT----- 318

Db 89 MetIleGluaspValGlnSerLeuLeuAspGluGluGlnGlnMetPheAlaPheArg 108

Qy 319 -----TCCAAAGACACAGTCAGCCCGAGCCTCCGATCGTACTATGTAACAG 366

Db 109 SerArgAlaArgSerThrAspThrPheAsn-----TyrAlaThr 121

Qy 367 TATCACTCACTAAATGAATCTATTCTTGATAGATTTTATACTGAGAGCATCTCGAT 426

Db 122 TyrHisThrLeuGluGluIleTyrAspPheLeuAspLeuValAlaGluAsnProHis 141

Qy 427 ATGCTTACAAAAATCCACATCTGATCTCATTTGAGAAGTACCACCTCTATGTTTAAAG 486

Db 142 LeuValSerLysIleGlnIleGlyAsnThrTyrGluGlyArgProIleTyrValLeuLys 161

Qy 487 GTTCTGGAAGAACAAACCAACCAAAATATGATATGATTCATGATTCGATTCATGCC 546

Db 162 PheSer---ThrGlyGlySerLysArgProAlaIleTyrPheAlaLysLysIleHisSer 180

Qy 547 AGAAGATGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 585

Db 181 ArgGluTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThrGlnAspTyr 200

Qy 585 ----- 585

Db 201 GlyClnAspAlaAlaPheThrAlaIleLeuAspThrLeuAspIlePheLeuGluIleVal 220

Qy 586 -----GGCCATAATCGAATGTGGAGAGAAACCGT 615

Db 221 ThrAsnProaspGlyPheAlaPheThrHisSerThrAsnArgMetTrpArgLysThrArg 240

Qy 616 TCTTTCTATCGGAACATCATTCATCGGACAGACCTGAATAGCACTTTGTTCTCAAA 675

Db 241 SerHisThrAlaGlySerLeuGlyValAspProAsnArgAsnTrp---AspAla 259

Qy 676 CACTGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTCTGCACTTAT 735

Db 260 GlyPheGlyLeuSerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyLysPhe 279

Qy 736 CTTGAGTCAAGAACAGAGTGAAGCAGTGGCTAGTTCTTGTGAGAAGAAATATCAACCAG 795

Db 280 AlaAsnSerGluValGluValLysSerIleValAspPheValLysAspHisGlyAsn--- 298

Qy 796 ATTAAAGACATACATCAGATGCTATCTATCTCCAGCATATAGTGTTCATATTCCTAT 855

Db 299 IleLysAlaPheIleSerIleHisSerTyrSerGlnLeuLeuMetTyrProTyrGlyTyr 318

Qy 856 ACAGGAAGTAAAGCAAGACCATGAGGAACCTCTCTAGTACCCAGTGAAGCAGTTCGT 915

Db 319 LysThrGluProValProaspGlnAspGluLeuAspGlnLeuSerLysAlaAlaValThr 338

Qy 916 GCTATTGACAAAATACTAGTAAAAATACACAGTATACATGCGCATGCGCTCAGAAACCTTA 975

Db 339 AlaLeu---AlaSerLeuTyrGlyThrLysPheAsnTyrGlySerIleIleLysAlaIle 357

Qy 976 TACTAGTCTCTGAGGTGGGAGCATTTGATCTATGATTTGGGCATCAATATTCGTTT 1035

Db 358 TyrGlnAlaSerGlySerThrIleAspTrpThrTyrSerGlnGlyIleLysTyrSerPhe 377

Qy 1036 ACA 1038

Db 378 Thr 378

RESULT 14

CBP4_HUMAN

ID CBP4_HUMAN STANDARD; PRT; 421 AA.

AC Q5UII2; 2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Carboxypeptidase A4 precursor (BC 3.4.17.-) (Carboxypeptidase A3).

GN CPA4 OR CPA3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RESULT 15

```

RESUL3 13
CBPZ SIMVI STANDARD; PRT; 304 AA.
AC P42788;
AD ID CBPZ SIMVI
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
DS Simulium vittatum (Black fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Gut;
RC MEDLINE=94093864; PubMed=8269093;
RX Ramos A., Mahowald A., Jacobs-Lorena M.;
RA "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins.";
RL Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: Involved in the digestion of the blood meal.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Gut-specific.
CC -1- INDUCTION: By blood meal.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L08481; AAA18531.1; --
CC HSSP: P48052; LAYE.
CC MEROPS; M14.0PA; -.
CC InterPro; IPR000834; Zn_carboPept.
CC Pfam; PF00246; Zn_carboPept; 1.
CC SMART; SM00631; Zn_pept; 1.
CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE; PS00133; CARBOXYPEPT_ZN2; 1.
CC Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
KW NON TER 1
FT METAL 58 58 ZINC (BY SIMILARITY).
FT FT METAL 61 61 ZINC (BY SIMILARITY).
FT FT METAL 184 184 ZINC (BY SIMILARITY).
FT FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
FT FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
FT FT DISULFID 125 148 BY SIMILARITY.
SQ SEQUENCE 304 AA; 334849 MW; 2E63FF8A6AA9144 CRC64;

```

US-09-980-881A-1 COPY 18 1097 (1-1080) x CBPZ SIMVI (1-304)

364	QY	CAGTATCACTCACTAAATGA	AATCATTTCTTGGATAGAA	TTTTATATACTGAGAGGCATCT	423
	Db				
1	GlnTyrHisThrLeuProGlu	IleTyrSerTrpLeuAspArg	LeuValGluGluHisPro	20	
424	QY	GATATGCTTTACAAAAATCC	ACATTTGGATCTCATTTGGA	AGATACCCACTCTATGTTT	483
	Db				
21	GluHisValGluProValGlu	IleTyrSerTyrGluGluArg	GluLeuArgGlyVal	40	
484	QY	AAGGTTTCTCGAAAAAGAA	CAAAACAGCCAAAAATGCC	ATATGGATTGACTGTGGAAT	543

```

41  LysValSerTyrLysLys-----GlyAsnProValValMetValGluSerAsnIleHis 58
544 GCACAGAAATGGGATCTCTCTCTGCT-----TTCTGCTTG----- 576
59  AlaArgGluTrpIleThrAlaAlaThrThrThrTyrLeuLeuAsnGluLeuLeuThrSer 78
577 -----TGGTTTCATA----- 585
79  LysAsnSerThrIleArgGluMetAlaGluAsnTyrAspTrpTyrIlePheProValThr 98
586 -----GGCCAT-----AATCGAATGTGGAGAAAGAACCGTTCT 618
99  AsnProAspGlyTyrValTyrThrHisThrThrAspArgMetTrpArgLysThrArgSer 118
619 TTCATCGCAACAATCATTCGATCGGAACAGACCTGAAATAGCAACCTTGTCTCCAAACAC 678
119 ProAsnProAspSerLeuCysAlaGlyThrAspProAsnArgAsnTrp---AsnPheHis 137
679 TGTGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACACTTACTGTGAGCTTTATCCCT 738
138 TrpMetGluGlnGlyThrSerSerArgProCysThrGluThrTyrGlyGlyLysLysAla 157
739 GAGTCAGAACCCAGAGTGAAGGCAGTGGCTAGTTCCTTGAGAGAAATATCAACCAAGATT 798
158 PheSerGluValGluThrArgSerPheSerAspPheLeuLysThrLeuLysGlyGlnIle 177
799 AAGCATCATCATCAGCATGCATTCATCATCTCCAGCATATAGTGTTCATATTCCTATACA 858
178 LysValTyrLeuAlaPheHisSerTyrSerGlnLeuLeuLeuPheProTyrGlyHisThr 197
859 CGAAGTAAAGCAAGCAACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCT 918
198 CysGlnHisThrTyrAsnHisAspAspLeuGlnAlaIleGlyAspAlaAlaArgSer 217
919 ATTGACAAAACCTAGTAAAAATACCAGGTATACACATGGCCCATGGCTCAGAAACCTTATAC 978
218 LeuAlaGln---ArgTyrGlyThrAspTyrThrValGlyAsnIleTyrAspAlaIleTyr 236
979 CTAGTCTCTGGAGGTGGGACGAGTTCGACTAT-----GATTGGGCATCAAAATAT 1029
237 ProAlaSerGlyGlySerMetAspTrpAlaIaTyrAspThrLeuAspIleProIleAlaTyr 256
1030 TCGTTTACATCAAAACCCA-----CCTGTGAGAGAAGCTT 1062
257 ThrTyrGluLeuArgProArgAspGlyTrpAsnGlyPheGlnLeuProAlaAsnGlnIle 276
1063 TTGCCCG 1068
277 IlePro 278

```

Search completed: January 1, 2004, 18:54:08
Job time : 30.3984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:10:36 ; Search time 54.1425 Seconds
(without alignments)
10294.941 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097

Perfect score: 1932

Sequence: 1 atgaagcttgcgccttgc.....tttgcgcgtctctctaaaa 1080

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135748_24954/app_query.fasta_1.2958
-DB=SPTRMBL_23 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORES=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN 1 1 190 @runat_31122003_135748_24954 -NCPU=6 -ICPU=3
-NO MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DBLOP=6 -DELEXT=7

Database :

SPTRMBL_23:*
1: sp archaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp invertebrate:*
6: sp mammal:*
7: sp mhc:*
8: sp organelle:*
9: sp phage:*
10: sp plant:*
11: sp rodent:*
12: sp virus:*
13: sp vertebrate:*
14: sp unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1884	97.5	360	4	Q9P2Y6 homo sapien

2	1835	95.0	423	4	Q15114
3	1830	94.7	423	4	Q961Y4
4	1550.5	80.3	422	11	Q9JHH6
5	1544.5	79.9	422	11	Q9QZF0
6	1522.5	78.8	422	11	Q9EQV9
7	1029	53.3	198	4	Q9NTI8
8	706	36.5	416	13	Q9PUF2
9	688.5	35.6	408	13	Q8AXN3
10	646	33.4	416	6	Q9XGP3
11	630.5	32.6	412	11	P97597
12	611.5	31.7	437	4	Q8N4T0
13	610	31.6	434	13	Q8AVS2
14	581	30.1	419	13	Q8UUK1
15	564.5	29.2	374	4	Q8IVL8
16	554.5	28.7	436	11	Q8R4H4
17	529.5	27.4	419	13	Q8AXN4
18	519.5	26.9	436	4	Q8WXQ8
19	516.5	26.7	419	6	Q9TV85
20	506.5	26.2	420	11	Q8BWK6
21	484.5	25.1	452	13	Q8UW45
22	483	25.0	262	11	Q8BVD0
23	472	24.4	247	4	Q96KZ9
24	460.5	23.8	419	13	Q8AXN5
25	460	23.8	424	5	Q9VL87
26	460	23.8	428	5	Q8IGC4
27	459	23.8	1132	5	Q9W475
28	451.5	23.4	422	5	Q9VLZ2
29	440.5	22.8	226	4	Q8TDE8
30	436	22.6	584	5	Q9TZH1
31	434.5	22.5	351	5	Q8TQJ1
32	434	22.5	351	4	Q8NA08
33	418	21.6	445	5	Q9W478
34	416	21.5	488	5	Q9BL88
35	412.5	21.4	416	5	Q9VRZ3
36	407.5	21.1	419	5	Q9VLZ1
37	402	20.8	424	5	Q9VS66
38	401.5	20.8	430	5	Q8SZP6
39	400.5	20.7	430	5	Q9VL86
40	399.5	20.7	427	5	Q9U9K2
41	397	20.5	323	5	Q19121
42	395.5	20.5	540	5	Q9TZH2
43	394	20.4	453	5	Q9V342
44	385	19.9	423	5	Q61532
45	382.5	19.8	581	5	Q23318

ALIGNMENTS

RESULT 1

ID	Q9P2Y6	PRELIMINARY;	PRT;	360 AA.
AC	Q9P2Y6;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Carboxypeptidase B-like protein.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Matsumoto A.;			
RT	"Isolation, molecular cloning, and partial characterization of a novel			
RT	carboxypeptidase B from human plasma.";			
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92042093; PubMed=1939207;			
RA	Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;			
RT	"Isolation, molecular cloning, and partial characterization of a novel			
RT	carboxypeptidase B from human plasma.";			
RL	J. Biol. Chem. 266:21833-21838(1991).			

```
DR EMBL; AB011969; BAA90475.1; -.
DR HSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR00834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_dept; 1.
DR SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

Alignment Scores:
Pred. No.: 7.91e-181 Length: 360
Score: 1884.00 Matches: 354
Percent Similarity: 98.99% Conservative: 2
Best Local Similarity: 98.33% Mismatches: 4
Query Match: 97.52% Indels: 0
DB: 4 Gaps: 0

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9P2Y6 (1-360)

Qy 1 ATGAAGCTTTCAGCGCTGAGTCTGTACCAATGTTCTCTCTGTGAGCAGCATGTC 60
Db 1 MetLysLeuCySerLeuAlaValLeuValProilleValLeuPheCysGluGlnHisVal 20
Qy 61 TTCGGTTCAGAGTGGCCAGTCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnThrGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGAACTTACTACACATATGAGATTTCTCTGGCAGCCGGTAAACAGCTGAC 180
Db 41 ValLeuGlnAenLeuThrThrTy:GluileValLeuTrpGlnProValThrAlaAsp 60
Qy 181 CTTATTGTGAAGAAAAACAAGTCCATTTTGTGTAATGATCATCTGATGTCGACAAATGTG 240
Db 61 LeuileValLysLysGlnValHisPhePheValAenAlaSerAspValAsnVal 80
Qy 241 AAAGCCATTTAAATGTGAGCGGAATCCAGCAGTCTCTGCTGGCAGCGTGGAGAGAT 300
Db 81 LysAlaHisLeuAenValSerGlyileProCySerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGATTTTCAACAGCAGACAGTCCAGCCCGAGCTCCGATCTCTAT 360
Db 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTy:Ty 120
Qy 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 420
Db 121 GluGlnTyHisSerLeuAenGluileTySerTrpilleGluPheileThrGluArgHis 140
Qy 421 CCTGATGCTTCAAAAATCCATGATGCTCTCTCATTTGAGAGTACCCACTCTATGTT 480
Db 141 ProaspMetLeuThrLysileHisileGlySerSerPheGluLysTyProLeuTyVal 160
Qy 481 TTAAGGTTTCTGGAAGAACAAACAGCCAAAATGCCATATGATGATGATGGAATC 540
Db 161 LeuLysValSerGlyLysGluGlnAlaLysAenAlalleTrpilleAspCysGlyile 180
Qy 541 CATGCCAGAGATGATCTCTCTGCTTTCTGCTTGTGTTGTTTCATAGGCCATAATCGAATG 600
Db 181 HisAlaargGluTrpilleSerProAlaPheCysLeuTrpPheilleGlyHisAenArgMet 200
Qy 601 TGGAGAAGACCCCTCTCTTCTATCGGACAAATCATTTGTCATCGGAACAGACTGAATAGC 660
Db 201 TrpArgLysAenArgSerPheTyAlaAenAenHisCysilleGlyThrAspLeuAenArg 220
Qy 661 AACCTTTGTCTCAAAACACTGTGTGAGGAAGGTGCATCCAGTCTCTCATGCTCGGAAC 720
Db 221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerGluThr 240
Qy 721 TACTGTGACATTTATCTGAGTCAAGAACAGAGTGAAGGAGTGGCTGATTTCTTGAGA 780
Db 241 TyCysGlyLeuTyProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
```

```
Qy 781 AGAAATATCAACAGAGATTAAGATATACATCAGATGATTCATATCTCCAGCATATAGTG 840
Db 261 ArgAsnileAenGlnileLysAlaTyrlleSerMethisSerTySerGlnHisileVal 280
Qy 841 TTTCCATATTTCTATACACGAAGTAAAGCAAGACCATGAGGAAGTCTCTCTAGTAGCC 900
Db 281 PheProTySerTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
Qy 901 AGTGAACAGTTCGTCTATTGACAAAACCTAGTAAAATACCAGGTATACATGCGCAT 960
Db 301 SerGluAlaValArgAlaileGluLysileSerLysAenThrArgTyThrHisGlyHis 320
Qy 961 GGCTCAGAAAACCTTATACCTAGCTCTGGAGGTGGGACCATGATGATCTATGATTTGGC 1020
Db 321 GlySerGluThrLeuTyLeuAlaProGlyGlyAspAspTrpilleTyAspLeuGly 340
Qy 1021 ATCAATATTTCTGTTTATCATCAAAACCCACCTGTAGAGAAGCTTTTCCGCTGTCTCTAAA 1080
Db 341 IleLysTySerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360

RESULT 2
Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PCPB protein.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSP; P00730; 1PTV.
DR MEROPS; M14.009; -.
DR Genew; HGNC:2300; CPB2.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR00834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_dept; 1.
FT CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6BE79F4 CRC64;

Alignment Scores:
Pred. No.: 7.06e-176 Length: 423
Score: 1835.00 Matches: 356
Percent Similarity: 86.23% Conservative: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 4 Gaps: 2

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q15114 (1-423)

Qy 1 ATGAAGCTTTCAGCGCTGAGTCTGTACCAATGTTCTCTCTGTGAGCAGCATGTC 60
Db 1 MetLysLeuCySerLeuAlaValLeuValProilleValLeuPheCysGluGlnHisVal 20
Qy 61 TTCGGTTCAGAGTGGCCAGTCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnThrGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGAACTTACTACACATATGAGATTTCTCTGGCAGCCGGTAAACAGCTGAC 180
```

Db 41 ValLeuGlnAsnLeuThrThrThrTyGluLeuValLeuTrpGlnProValThrAlaAsp 60
181 CTTATTGTGAAGAAAAACAAGTCATTTTGTAAATGCATCTGATGTCGACAAATGTG 240
Db 61 LeuileVallyLysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 80
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATCCATGCAGTGTCTTCTGGCGAGAGCTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGATTCCAAACACACACAGTCCAGCCCGAGCTCCGATCTGATCTAT 360
Db 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyTyr 120
Qy 361 GAACAGTATCTACTCAATAATGAAATCTATTCTTGGATGAGATTTATAACTGAGAGCAT 420
Db 121 GluGlnTyHisSerLeuAsnGluLeuTySerTrpIleGluPheIleThrGluArgHis 140
Qy 421 CCTGATATGCTTACAAAATCCATTTGGATCCCTCATTTGAGAGTACCCGACTCTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeuTyVal 160
Qy 481 TTAAGGTTTCTGGAAGAAACAACAGCAGCAAAATGCCATATGATGATGCTGGAATC 540
Db 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Qy 541 CATCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGGCCAT----- 591
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
Qy 591 ----- 591
Db 201 PheTyGlyIleIleGlyGlnTyThrAsnLeuLeuArgLeuValAspPheTyValMet 220
Qy 592 -----NATCGAATGCGAGAAG 609
Db 221 ProValValAsnValAspGlyTyAspTySerTrpLysLysAsnArgMetTrpArgLys 240
Qy 610 AACGGTCTTTCTATGCGAACAATCATTTGATCGGAACAGACACCTGGAATAGCAATTTGTC 669
Db 241 AsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
Qy 670 TCCAAACACTGTTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGAAACCTACTGTGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyCysGly 280
Qy 730 CTTTATCTGAGTACAGACAGAGTGAAGCAGTGGCTAGTTTCTTGGAGAGAAATATC 789
Db 281 LeuTyProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 300
Qy 790 AACAGATTAAGCATATACATCAGCATGCATTCTACTCCAGCATATAGTGTTCATAT 849
Db 301 AsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleValPheProTy 320
Qy 850 TCCTATACAGAACTAAAGCAAGACATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 909
Db 321 SerTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
Qy 910 GTTGTGCTATTGACAAACTAGTAAATACAGGTATACACATGCGCCATGGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisGlyHisGlySerGlu 360
Qy 970 ACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGATGATTTGGGCATCAAAATAT 1029
Db 361 ThrLeuTyLeuAlaProGlyGlyLysAspTrpIleTyAspLeuGlyIleLysTy 380
Qy 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyGlyPheLeuLeuProGluArgTy 400
Qy 1038 ATCAAAACCCACTGTAGAGAGCTTTTGGCGCTGCTCTAAA 1079

Db 401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414
Qy 961Y4
ID Q961Y4 PRELIMINARY; PRT; 423 AA.
AC Q961Y4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B2 (Plasma).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007057; AA007057.1; -.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn carbOpept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_dept; 1.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6EE7CF5 CRC64;
Alignment Scores:
Pred. No.: 2.25e-175 Length: 423
Score: 1830.00 Matches: 355
Percent Similarity: 85.99% Conservative: 1
Best Local Similarity: 85.75% Mismatches: 4
Query Match: 94.72% Indels: 55
Gaps: 2
US-09-980-881a-1_copy_18_1097 (1-1080) x Q961Y4 (1-423)
Qy 1 ATGAAGCTTTTGACGCTTGTGAGTCTTGTGATCCCATTTCTTCTTCTGTGAGCAGCATGTC 60
Db 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Qy 61 TTCGGTTCAGAGTGCCCAAGTTCTAGCTCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGAACTTACTACAAATATGAGATTTCTTCTGCGAGCGGTAACAGCTGAC 180
Db 41 ValLeuGlnAsnLeuThrThrTyGluIleValLeuTrpGlnProValThrAlaAsp 60
Qy 181 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 240
Db 61 LeuileVallyLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATTTCCATGAGTGTCTTGTGGCAGAGCTGGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGATTTCCAAACGACACAGTCCAGCCCGAGCTCCGATCTGATCTAT 360
Db 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyTyr 120
Qy 361 GAACAGTATCTACTCAATAATGAAATCTATTCTTGGATGAGATTTATAACTGAGAGCAT 420
Db 121 GluGlnTyHisSerLeuAsnGluLeuTySerTrpIleGluPheIleThrGluArgHis 140
Qy 421 CCTGATATGCTTACAAAATCCACATTTGGATCTCTTCTAGAGTACCCGACTCTATGTT 480
Db 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeuTyVal 160
Qy 481 TTAAGGTTTCTGGAAGAAACAACAGCAGCAAAATGCCATATGATGATTTGGAATC 540

```
Db 161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrrPileAspCysGlyIle 180
Qy 541 CATCCAGAGATGATCTCTCCTGCTTCTGCTTGGTTTCATAGCCAT----- 591
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200
Qy 591 ----- 591
Db 201 PheTyrglylleIleGlyGlnTyThrAsnLeuLeuArgLeuValAspPheTyrValMet 220
Qy 592 -----AATCGAATGTGAGAAAG 609
Db 221 ProValValAsnValAspGlyTyAspTySerTrpLysLysAsnArgMetTrpArgLys 240
Qy 610 AACGGTCTTCTATGCGAACAATCATTTGATCGGAACAGCTGAATAGCACTTTGTC 669
Db 241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260
Qy 670 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATGCTCGGAACCTACTGTGGA 729
Db 261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrcysgly 280
Qy 730 CTTTATCTCTGAGTACAGAACAGAGTGAAGCAGTGGCTAGTTCTTCTGAGAAGAAATATC 789
Db 281 LeuTyrrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300
Qy 790 AACAGATTAAAGATATACATCAGATGATCATCTATCTCCAGCATATAGTTTCCATAT 849
Db 301 AsnGlnIleLysAlaTyrlleSerMetHisSerTyrrSerGlnHisIleValPheProTy 320
Qy 850 TCCTATACAGCAAGTAAAGCAAGACCATGAGCAAGTCTCTAGTAGCAGTGAAGCA 909
Db 321 SerTyrrThrArgSerLysSerLysPheGluGluLeuSerLeuValAlaSerGluAla 340
Qy 910 GTTCGTCTATTGACAAACTAGTAAATAACACAGGTATATACATGCGCCATGGCTCAGAA 969
Db 341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrrThrHisGlyHisGlySerGlu 360
Qy 970 ACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAATAT 1029
Db 361 ThrLeuTyrrLeuAlaProGlyGlyGlyAspAspTrpIleTyrrAspLeuGlyIleLysTy 380
Qy 1030 TCGTT-----TAC 1037
Db 381 SerPheThrIleGluLeuArgAspThrGlyThrTyrrGlyPheLeuLeuProGluArgTy 400
Qy 1038 ATCAAAACCCCTGTAGAGAGCTTTTGGCGCTGTCTCTAAA 1079
Db 401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414
RESULT 4
Q9JHH6 PRELIMINARY; PRT; 422 AA.
ID AC Q9JHH6
DC Q9JHH6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE (1110032P04Rik protein)
DE CPB2 OR TAFI OR 1110032P04RIK.
GN Mus musculus (Mouse).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T.; Miwa T.; Akatsu H.; Matsukawa N.; Obata K.; Okada N.;
RA Campbell W.; Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RL whereas carboxypeptidase R is not.";
RL J. Immunol. 165:1053-1058(2000).
RN (2)
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=20201986; PubMed=10739389;
RA Marx P.F.; Wagenaar G.T.M.; Reijerkerk A.; Tiekstra M.J.;
RA van Rossum A.G.S.H.; Gebbink M.F.G.B.; Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
inhibitor.";
RL Thromb. Haemost. 83:297-303(2000).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J.; Shinagawa A.; Shibata K.; Yoshino M.; Itoh M.; Ishii Y.;
RA Arakawa T.; Hara A.; Fukunishi Y.; Konno H.; Adachi J.; Fukuda S.;
RA Aizawa K.; Izawa M.; Nishi K.; Kiyosawa H.; Kondo S.; Yamanaka I.;
RA Saito T.; Okazaki Y.; Gojobori T.; Bono H.; Kasukawa T.; Saito R.;
RA Kadoya K.; Matsuda H.A.; Ashburner M.; Batalov S.; Casavant T.;
RA Fleischnann W.; Gaasterland T.; Gissi C.; King B.; Kochiwa H.;
RA Kuehl P.; Lewis S.; Matsuo Y.; Nikaido I.; Pesole G.; Quackenbush J.;
RA Schriml L.M.; Staubli P.; Suzuki R.; Tomita M.; Wagner L.; Washio T.;
RA Sakai K.; Okido T.; Furuno M.; Aono H.; Baldarelli R.; Barah G.;
RA Blake J.; Boffelli D.; Bojunga N.; Carninci P.; de Bonaldo M.F.;
RA Brownstein M.J.; Bult C.; Fletcher C.; Fujita M.; Gariboldi M.;
RA Gustincich S.; Hill D.; Hofmann M.; Hume D.A.; Kamiya M.; Lee N.H.;
RA Lyons P.; Marchionni L.; Mashima J.; Mazzarelli J.; Mombaerts P.;
RA Nordone P.; Ring B.; Ringwald M.; Rodriguez I.; Sakamoto N.;
RA Sasaki H.; Sato K.; Schoenbach C.; Seva T.; Shibata Y.; Storch K.-F.;
RA Wynshaw-Boris A.; Wang K.H.; Weitz C.; Whittaker C.; Wilming L.;
RA Wynshaw-Boris A.; Yoshida K.; Hasegawa Y.; Kawaji H.; Kohseki S.;
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AB021968; BAB03402.1; -
DR EMBL; AF164524; AAF62385.1; -
DR EMBL; AK004045; BAB23141.1; -
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_dept; 1.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D555CB CRC64;
Alignment Scores:
Pred. No.: 3,2e-147 Length: 422
Score: 1550.50 Matches: 302
Percent Similarity: 78.99% Conservative: 25
Best Local Similarity: 72.95% Mismatches: 32
Query Match: 80.25% Indels: 56
DB: 11 Gaps: 3
US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9JHH6 (1-422)
Qy 1 ATGAGCTTTGCAGCCTTGAGTCCCTTGATCCCATTTCTCTCTCTGTGAGCAGCATGTC 60
Db 1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleLeuTyrr---GluGlnHisGly 19
Qy 61 TTCGCGTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
Db 20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39
Qy 121 GTTCTACAGAAATCTTACTACAAATATGAGATTGTCTCTGGCAGCGCGGTAAACAGCTGAC 180
Db 40 LeuLeuGlnAsnLeuThrThrTyrrGluValValLeuTrpGlnProValThrAlaGlu 59
Qy 181 CTTATTCTGAAGAAAACAAGTCCCATTTTTTTGTAATGCATCTGATGTCGCAATGTG 240
Db 60 PheIleGlnLysLysGlyGlnValHisPhePheValAsnAlaSerAspValAspSerVal 79
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTCTTGTGTCGAGACGTCGAAGAT 300
```

Db 80 LysAlaHisLeuAenValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99
Qy 301 CTTATTCAACAGCAGATTCCACAGACACAGTACGCCCCGAGCCTCCGATCCTACTAT 360
Db 100 LeuIleGluGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyTy 119
Qy 361 GAACAGTATCACTCACTAATGAATCTATTCTTGGATAGATTATTAACAGAGCAT 420
Db 120 GluGlnTyHisSerLeuAsnGluIleTySerTrpIleGluValIleThrGluGlnHis 139
Qy 421 CCTGATATGCTTACAAAATCCACATTGGATTCCTCATTTGAGAAAGTACCCACTCTATGTT 480
Db 140 ProAspMetLeuGlnLysIleTyIleGlySerSerPheGluLysTyProLeuTyVal 159
Qy 481 TTAAGGTTTCTGAAAGAAACAAACAGCCAAAATGCCATATGATGATGATGGAATC 540
Db 160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179
Qy 541 CATGCCAGAGATGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 591
Db 180 HisAlaArgGlnTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyValThrGln 199
Qy 591 ----- 591
Db 200 PheHisGlyLysGluAsnLeuTyThrArgLeuLeuArgHisValAspPheTyIleMet 219
Qy 592 -----AATCGAATGTGGAGAAAG 609
Db 220 ProValMetAsnValAspGlyTyAspTyThrTrpLysLysAsnArgMetTrpArgLys 239
Qy 610 AACCGTTCTTCTATCGGAACAACATTCATCGCATCGGAACAGACCTGAATAGCACTTTGTC 669
Db 240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259
Qy 670 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATCTCGGAACCTACTGGA 729
Db 260 SerLysHisTrpCysGluLysGlyAlaSerSerSerSerCysSerGluThrTyCysGly 279
Qy 730 CTTTATCCTAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTCTGAGAGAAATATC 789
Db 280 LeuTyProGluSerGluProGluValLysAlaValAlaAspPheLeuArgAsnIle 299
Qy 790 AACAGATTAAAGCATATCATCATGATTCATCTCCGAGCATATAGTTTCCATAT 849
Db 300 AspHisIleLysAlaTyIleSerMetHisSerTySerGlnGlnIleLeuPheProTy 319
Qy 850 TCCTATACCAAGTAAAGCAAGACCATCAGCAACTGCTCTAGTAGCCAGTGAAGCA 909
Db 320 SerTyAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339
Qy 910 GTTCGTGCTATTGACAAAACCTAGTAAATAACAGGTATACACATGGCCATGGCTCAGAA 969
Db 340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyThrHisGlySerGlySerGlu 359
Qy 970 ACCTTATACCTAGCTCTGAGAGTGGGACGATTGGATCTATGATTTGGGCATCAATAT 1029
Db 360 SerLeuTyLeuAlaProGlyGlySerAspSerTrpIleTyAspLeuGlyIleLysTy 379
Qy 1030 TCGTT-----TAC 1037
Db 380 SerPheThrIleGluLeuArgAspThrGlyArgTyGlyPheLeuLeuProGluArgTy 399
Qy 1038 ATCAACCCACTGTAGAGAAGCTTTTGCCTGCTCTCTAAA 1079
Db 400 IleLysProThrCysAlaGluAlaLeuAlaIleSerLys 413

RESULT 5
Q9QZF0 ID Q9QZF0 PRELIMINARY; PRT; 422 AA.
AC Q9QZF0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Carboxypeptidase U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J;
RA He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
gene";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF186188; AAF00528.1; -;
DR HSP; P00730; 5CPA.
DR MEROPS; M14.009; -;
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 422 AA; 48832 MW; B53PFFB09943954E CRC64;
Alignment Scores:
Pred. No.: 1,29e-146 Length: 422
Score: 1544.50 Matches: 301
Percent Similarity: 78.74% Conservative: 25
Best Local Similarity: 72.71% Mismatches: 33
Query Match: 79.94% Indels: 56
DB: 11 Gaps: 3
US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9QZF0 (1-422)
Qy 1 ATGAAGCTTTGCGAGCCTTGAGTCTTGTACCCATCTTCTCTCTGAGCAGCATGTC 60
Db 1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleLeuTyfyr---GluGlnHisGly 19
Qy 61 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Db 20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39
Qy 121 GTTCTACAGAACTTACTACAACATATGAGATTGTTCTTGGCAGCCGGTAACAGTGCAC 180
Db 40 LeuLeuGlnAsnLeuThrThrTyThrTyGluValValLeuTrpGlnProValThrAlaGlu 59
Qy 181 CTTATTGTGAGAAAAAACAAGTCCATTTTGTAAATGATGATGATGATGATGATGATG 240
Db 60 PheIleGluLysLysLysGluValHisPheValAsnAlaSerAspValAspSerVal 79
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATCCCATGAGTGTCTTGTGGCAGCAGTGGAAAGAT 300
Db 80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99
Qy 301 CTTATTCAACAGCAGATTTCACACGACACAGTACGCCCCGAGCCTCCGATCCTACTAT 360
Db 100 LeuIleGluGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyTy 119
Qy 361 GAACAGTATCACTCACTAATGAATCTATTCTTGGATAGATTATTAACAGAGCAT 420
Db 120 GluGlnTyHisSerLeuAsnGluIleTySerTrpIleGluValLeuThrGluGlnHis 139
Qy 421 CCTGATATGCTTACAAAATCCACATTGGATTCCTCATTTGAGAAAGTACCCACTCTATGTT 480
Db 140 ProAspMetLeuGlnLysIleTyIleGlySerSerPheGluLysTyProLeuTyVal 159
Qy 481 TTAAGGTTTCTGAAAGAAACAAACAGCCAAAATGCCATATGATGATGATGGAATC 540
Db 160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179
Qy 541 CATGCCAGAGATGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 591


```

Qy 790 AACGAGATTAAAGCATACATCAGCATGATTCATCTCCAGCATATAGTGTTCATAT 849
Db 300 AsnHisileysAlaTyrlleSerMetHisSerTyrlSerGlnGlnlleLeuPheProTyrl
Qy 850 TCCTATACACGAAGTAAAGAGACATCAGAGAACTGCTCTAGTAGCCAGTGAAGCA 909
Db 320 SerTyrlSerArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAla 339
Qy 910 GTTCGTCTATTGCAAAACCTAGTAAATAACACAGGTATACACATGCCATGGCTCAGAA 969
Db 340 ValArgAlalleGluSerlleAsnLysAenThrArgTyrlThrHisGlySerGlySerGlu 359
Qy 970 ACCTTATACCTAGTCTCTGGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATAT 1029
Db 360 SerLeuTyrlleuAlaProGlyGlySerAspTrpIleTyrlAspLeuGlylleLysTyrl 379
Qy 1030 TCGTTTATCATC----- 1040
Db 380 SerPheThrilleGluLeuArgAspThrGlyArgTyrlGlyPheLeuLeuProGluArgPhe 399
Qy 1041 ---AAACCCACCTGTAGAGAGCTTTTGGCCGCTGTCTCTAAA 1079
Db 400 IleLysProThrCysAlaGluAlaLeuAlaValSerLys 413

RESULT 7
Q9NTI8 PRELIMINARY; PRT; 198 AA.
AC Q9NTI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BAI39H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL337141; CAB92622.1; -.
DR HSSP; P00730; 1CPX.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep M14.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
FT NON TER 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

Alignment Scores:
Pred. No.: 9.22e-95 Length: 198
Score: 1029.00 Matches: 196
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 53.26% Indels: 0
DB: 4 Gaps: 0

US-09-980-881a-1_COPY_18_1097 (1-1080) x Q9NTI8 (1-198)

Qy 1 ATGAGCTTTGCAGCTTGAGCTGCTGATCCCACTGCTCTCTCTGAGCAGCATGTC 60
Db 1 MetLysLeuCysSerLeuAlaValleuValProilleValleuPheCysGluGlnHisVal 20
Qy 61 TTCCGCTTCCAGAGTGGCAGAGTCTAGCTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Db 21 PheAlaPheGlnSerGlyGlnValleuAlaLeuProArgThrSerArgGlnValGln 40
Qy 121 GTTCTACAGAATCTTACTACAAATATAGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

```

```

Db 41 ValLeuGlnAsnLeuThrThrThrTyrlGluilleValleuTrpGlnProvalThrAlaAsp 60
Qy 181 CTTATTGTGAAGAAAAACAAAGTCCATTTTTTTGTAATGCATCTCATGTGCGCAATGTG 240
Db 61 LeuilleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTCTGGCAGAGCGTGAAGAT 300
Db 81 LysAlaHisLeuAsnValSerGlylleProCysSerValleuLeuAlaAspValGluAsp 100
Qy 301 CTTATTCAACAGCAGAGATTTCCAAACGACACAGTCCAGCCCGAGCCTCCGATCGTACTAT 360
Db 101 LeuileGlnGlnlleSerAsnAspThrValSerProArgAlaSerAlaSerTyrlTyrl 120
Qy 361 GAACAGTATCACCTACCTAAATGAAATCTATTCTTGGATAGATTTTATAACTGAGAGCAT 420
Db 121 GluGlnTyrlHisSerLeuAsnGluilleTyrlSerTrpIleGluPheilleThrGluArgHis 140
Qy 421 CTTGATATGCTTCAAAAATCCACATTTGGATTCCTCATTTGAGAAGTACCACCTATGTT 480
Db 141 ProAspMetLeuThrLyslleHislleGlySerSerPheGluLysTyrlProLeuTyrlVal 160
Qy 481 TTAAGGTTTCTGGAAGAGACAAACAGCCAAAATGCCATATGGATTTGACTGTGGATC 540
Db 161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlylle 180
Qy 541 CATGCCAGAGATGGATCTCTCTGCTTCTGCTGTGTGGTTTCATAGGCCAT 591
Db 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheilleGlyHis 197

RESULT 8
Q9PUF2 PRELIMINARY; PRT; 416 AA.
AC Q9PUF2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase homolog.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA Murbach A.F., Hayaishi M.A.F., Camargo A.C.M.;
RT "Screening of Bothrops jararaca pancreas cDNA library.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190274; AAF01344.1; -.
DR HSSP; P09955; INSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;

Alignment Scores:
Pred. No.: 3.71e-62 Length: 416
Score: 706.00 Matches: 149
Percent Similarity: 53.77% Conservative: 65
Best Local Similarity: 37.44% Mismatches: 124
Query Match: 36.54% Indels: 61
DB: 13 Gaps: 6

US-09-980-881a-1_COPY_18_1097 (1-1080) x Q9PUF2 (1-416)

Qy 43 TTCTGTGAGCAGCATGCTTTCGCGTTCAGAGTGGCCAGGTCTCTAGCTCTCTCTCTAG 102

```

```
Db 14 PheAlaGluThrThrValHieArgPheAspGlyGluLysValTyrArgValThrProArg 33
Qy 103 ACCTCTAGGCAAGTTCAGTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGG 162
Db 34 AsnGluAspGluValTyrPheLeuAsnTyrPheLeuAlaAsnIleValGlnValAspPheTrp 53
Qy 163 CAGCGGTAAACAGCTGACCTTATTGTGAAGAAAAACAAGTCCATTTTTTGTAAATGCA 222
Db 54 ArgProAspSerValGluLeuValLysAlaGluMetThrValAspPheArgIleGluAla 73
Qy 223 TCTGATCTCGACAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTGTCTTG 282
Db 74 AspArgCysSerGluValGluSerIleLeuGlnSerGlyLeuAsnTyrGluIleLeu 93
Qy 283 CTGCAGACGTGGAAGATCTTATTCAACAGACAGATTTCACAGCACAGTCCAGCCCGCA 342
Db 94 IleAspAsnLeuGlnAlaValLeuAspArgGlnLeuAspAsnHisAla-----Arg 110
Qy 343 GCCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTCTTGATAGAA 402
Db 111 ThrAlaGlyTyrAsnTyrGluLysTyrAsnSerTrpGluLysIleAspAlaTrpThrAla 130
Qy 403 TTTATACTGAGGCGCATCTGATATGCTTACAAAAATCCACATTTGGATCCTCATTTGAG 462
Db 131 AspIleAlaAsnGluAsnProSerLeuValSerArgLeuGlnIleGlyThrThrPheGlu 150
Qy 463 AAGTACCCTCTATGTTTAAAGGTTCTGGAAAAAGAACAAACAGCCCAAAATGCCATA 522
Db 151 GlyArgProMetProLeuLeuValLysValLysProGlyValAlaAsnLysLysAlaIle 169
Qy 523 TGGATTGACTGTGAATCCATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGCTTC 582
Db 170 PheIleAspCysGlyPheHieAlaArgGluTrpIleSerProAlaPheCysGlnTrpPhe 189
Qy 583 ATA----- 585
Db 190 ValArgGluAlaValArgThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnLys 209
Qy 586 -----GGCCAT 591
Db 210 LeuAspPheTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSerTrpLysGln 229
Qy 592 AATCGAATGTGGAGAAGACCGTTCTTATCGGAACATATTCATCGGAACAGAC 651
Db 230 SerArgMetTrpArgLysThrArgSerValAsnAlaGlySerThrCysIleGlyThrAsp 249
Qy 652 CTGAATACCACTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 711
Db 250 ProAsnArgAsnPhe---AspAlaAlaTrpCysSerValGlyAlaSerArgAsnProCys 268
Qy 712 TCGGAACCTCTGAGACTTATCTGAGTCAGACAGCAAGTGAAGCGAGTGGCTAGT 771
Db 269 SerGluThrTyrCysGlySerLysProGluSerGlyLysGluThrLysAlaLeuAlaAsp 288
Qy 772 TTCTTGAGAGAAATATCAACACGATTAAACATACATACATCAGCATTCATATCCCGAG 831
Db 289 PheIleArgArgAsnArgSerIleIleGlnAlaTyrLeuThrIleHisSerTyrSerGln 308
Qy 832 CATATAGTGTTCCTATATCTTATACAGAAAGTAAAGCAACCATGAGGAAGTCTCT 891
Db 309 MetLeuLeuTyrProTyrSerTyrThrTyrAspLeuThrSerAsnAsnLysLeuAsn 328
Qy 892 CTAGTAGCAGTGAAGAGTTCTGCTATTGACAAAACACTAGTAAATAATACAGATATACA 951
Db 329 SerIleAlaLysGluAlaIleArgGluLeu---LysValLeuPheGlyThrGluTyrThr 347
Qy 952 CATGCCATGCTCAGAAACCTTATACCTAGCTCTCTGAGGTGGGACGATGGATCTAT 1011
Db 348 TyrGlyProGlyAlaAlaThrIleTyrProAlaAlaGlyLysArgAspTrpAlaTyr 367
Qy 1012 GATTGGGCATCAATATTCGTTTACATC----- 1040
```

```
Db 368 AspGlnGlyIleLysTyrAlaPheThrPheGluLeuArgAspLysGlyArgTyrGlyPhe 387
Qy 1041 -----AAACCACCTGTAGAGAACCTTTTGGCCGCTGC 1073
Db 388 AlaLeuProGluSerGlnIleLysProThrCysGluGluThrMetIleAlaVal 405
RESULT 9
Q8AXN3 PRELIMINARY; PRT; 408 AA.
AC Q8AXN3
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE Carboxypeptidase B (Fragment).
GN CFB.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralicththyidae; Paralicthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RA Srivastava A.S., Kurokawa T., Suzuki T.;
RT "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT A2 and B from the Japanese flounder, Paralicthys olivaceus.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB099302; BAC53789.1; -.
FT NON TER 1
SQ SEQUENCE 408 AA; 45904 MW; 809CAACD9D2DA34B CRC64;
Alignment Scores:
Pred. No.: 2,14e-60 Length: 408
Score: 688.50 Matches: 149
Percent Similarity: 53.85% Conservative: 61
Best Local Similarity: 38.21% Mismatches: 121
Query Match: 35.64% Indels: 60
DB: Gaps: 6
US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8AXN3 (1-408)
```

```
Qy 67 TTCACAGAGTCCCAAGTTCTAGCTGCTTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db 12 PheGluGlyGluLysValPheArgLeuLysProValPheAspGluHisValThrLeuIle 31
Qy 127 CAGAACTTACTACACATATGAGATTGTTCTTGGCAGCGCGTAAACAGCTGACCTTATT 186
Db 32 ArgAspLeuAlaAsnSerIleGluValAspPheTrpArgProGluSerSerGlnLeuVal 51
Qy 187 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATCGACATGTGAAAGCC 246
Db 52 ThrIleAspIleAspValAspIleArgValProAlaIleTyrLeuAspMetValTyrThr 71
Qy 247 CATTTAAATGTAGCGGAATTCATGCGATGCTTGTGCGACACGTGGAAGATCTTATT 306
Db 72 LeuLeuGlnGlnSerAspMetGluHisGluValLeuIleGluAspLeuGlnSerAlaVal 91
Qy 307 CAACAGCAGATTTCCAAACGACAGCAGTCAGCCCCCGACCTCCCGATCGTACTATGAACAG 366
Db 92 AspAlaGluAlaAsp-----LeuLysProSerProArgAlaHisSerTyrThrLys 108
Qy 367 TATCACTCACTAAATGAATCTATTCTTGGATAGAATTTATTAAGTACGAGGATCTTGAT 426
Db 109 TyrAsnAsnLeuAspLysValGlnSerTrpIleAlaSerIleSerSerSerAsnProAsp 128
Qy 427 ATGCTTACAAAAATCCACATTTGGATCTCTATTTGAGAAGTACCCTCTATGTTTAAAG 486
Db 129 LeuIleSerLysGlnValIleGlyAsnThrTyrGluLysArgProMetThrValLeuLys 148
Qy 487 GTTTCGAAAAAGAACAAACAGCCAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
Db 149 LeuGlyLysLysSerSerThrLysProAlaIlePheMetAspCysGlyIleHieAla 168
```

```
QY 547 AGAGATGGATCTCTCTGCTTTCTGCTTGGTTTCATA----- 585
Db 169 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValLysGluAlaLeuSerThrTyr 188
QY 585 ----- 585
Db 189 GlySerAspSerGlnMetThrSerLeuLeuAsnGlnMetAspValPheValLeuProVal 208
QY 586 -----GQCAT-----AATCGAATGTGGAGAAACCGT 615
Db 209 PheAsnIleAspGlyTyrAspPheThrHisLysSerAsnArgMetTrpArgLysThrArg 228
QY 616 TCTTTCTATGCGAACAATCATTGTCAGACAGACCTGATAGCAATGCTTGTCTCCAAA 675
Db 229 SerArgLysSerGlySerSerCysIleGlyAlaAspProAsnArgAsnPhe---AspAla 247
QY 676 CACTGGTGTGAGGAAGTGATCCAGTCTCTCATGCTCGGAACCTACTGTGGGACTTTAT 735
Db 248 GlyTrpCysSerMetGlyAlaSerAspAsnProCysSerAspThrPheCysGlyTyrThr 267
QY 736 CTTGAGTCAGAACGAGTGAAGGAGTGTCTAGTCTTCTTGTGAGAGAAATATCAACAG 795
Db 268 ProGluSerGluIleGluValLysAsnValAlaAspPheIleArgAsnArgSerIle 287
QY 796 ATTAAAGCATACATCAGCATCATCTACTCCAGCATATAGTGTTCATATTCCTAT 855
Db 288 IleLysAlaTyrLeuThrValHisSerTyrSerGlnLeuLeuPheProTyrSerTyr 307
QY 856 ACAGGAAGTAAAGCAAGACCATGAGGAGCTGCTCTAGTAGCCAGTGAAGCAGTTCGT 915
Db 308 ThrTyrGlnLeuAlaAlaAspHisSerGluLeuMetLysValAlaGluGlyAlaSerAla 327
QY 916 GCTATTGACAAAATAGTAAAAATACCAGGTATACATGCGCCATGGCTCAGAAACCTTA 975
Db 328 AlaLeu---ArgSerLeuTyrGlyThrThrTyrThrSerGlyProGlyAlaThrIle 346
QY 976 TACCTAGCTCTGAGGTGGGAGCATGGATCATGATTTGGCATCAATATTCGTT 1035
Db 347 TyrProAlaAlaGlyGlySerAspTrpAlaTyrAspLeuGlyValLysTyrSerTyr 366
QY 1036 ACATC-----AAA 1043
Db 367 ThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleLys 386
QY 1044 CCCACCTGTAGAGAGCTTTTGGCGCTGTC 1073
Db 387 ProThrCysGluGluThrMetLeuAlaVal 396

RESULT 10
Q9XSP3
ID AC Q9XSP3 PRELIMINARY; PRT; 416 AA.
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Procarboxypeptidase B precursor (EC 3.4.17.2).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
ON NCBI_TaxID=9823;
RX [1]
SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=9321926; PubMed=10391940;
RA Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA Hershberger C., Aviles F.;
RT "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL J. Biol. Chem. 274:19925-19933 (1999).
DR EMBL; AJ133775; CAB46991.1; --
DR HSP; P03955; IPEA.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carboxypept.
```

```
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_dept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Hydrolase; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 416 PROCARBOXYPEPTIDASE B.
SQ SEQUENCE 416 AA; 47381 MW; 5FA7PCED7B45AE6A CRC64;

Alignment Scores:
Pred. No.: 4,1e-56 Length: 416
Score: 646.00 Matches: 140
Percent Similarity: 53.40% Conservative: 64
Best Local Similarity: 33.46% Mismatches: 134
Query Match: 36.65% Indels: 44
DB: 6 Gaps: 7

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9XSP3 (1-416)
QY 7 CTTTGACAGCTTGACAGCTTGTACCCATGTTCTTCTGTGAGCAGCATGTCTTCGCG 66
Db 1 MetLeuAlaPheLeuIleLeuValThrValThrLeuAlaSerAlaHisSerGlyGlu 20
QY 67 ---TTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTTAGSCAAAGTTCAAAGT 123
Db 21 HisPheGluGlyGlyLysValPheArgValAsnValGluAspGluAsnAspIleSerLeu 40
QY 124 CTACAGATCTTACTCAACATATGAGATTGTTCTCTGGCAGCGCGTGAACAGTGACCTT 183
Db 41 LeuHisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGln 60
QY 184 ATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTCATGTCGACGCAATGTGAAA 243
Db 61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGlu 80
QY 244 GCCCATTTAAATGTAGCGGAATTCATGCAAGTGTCTTGTGGCAGACGTGGAGATCTT 303
Db 81 AspPheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnLeuAsnLeuArgSerVal 100
QY 304 ATTCAACAGCAGATTTCCAAACAGACACAGTCAGCCCGACCTCCGATCGTACTATGAA 363
Db 101 LeuGluAlaGlnPheAspSerArgVal-----ArgThrThrGlyHisSerTyrGlu 117
QY 364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATACTGAGAGCATCTCT 423
Db 118 LysTyrAsnAsnTrpGluThrIleGluAlaTrpThrLysGlnValThrSerGluAsnPro 137
QY 424 GATATGCTTACAAAAATCCACATTTGGATCTCTATTGAGAAAGTACCCACTCTATGTTTA 483
Db 138 AspLeuIleSerArgThrAlaIleGlyThrThrPheLeuGlyAsnAsnIleTyrLeuLeu 157
QY 484 AAGCTTTCTGGAAGAACAAACAGCCAAAATGCCATATGGATGGATGATCTGTGAATCCAT 543
Db 158 LysVal---GlyLysProGlyProAsnLysProAlaIlePheMetAspCysGlyPheHis 176
QY 544 GCCAGAGAATGGATCTCTCTGCTTCTTCTGCTTCTGCTGTTTCATA----- 585
Db 177 AlaArgGluTrpIleSerHisAlaPheCysGlnTrpPheValArgGluAlaValLeuThr 196
QY 585 ----- 585
Db 197 TyrGlyTyrGluSerHisMetThrGluPheLeuAsnLysLeuAspPheTyrValLeuPro 216
QY 586 -----GQCAT-----AATCGAATGTGGAGAAAGAAC 612
Db 217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThr 236
QY 613 CGTTCTTTCTATCGGAACATCATTCATTCGTCGGAACAGACCTGATAGCACTTTGTCTCC 672
Db 237 ArgSerThrAsnAlaGlyThrThrCysIleGlyThrAspProAsnArgAsnPhe---Asp 255
```



```
Db 351 TyrIysThrSerGlySerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThrPhe 370
Qy 1036 AC-----ATCAAA 1043
Db 371 AlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLys 390
Qy 1044 CCACCTGTAGAGAGCTTTTGGCGCTGTC 1073
Db 391 ProThrCysLysGluThrMetLeuSerVal 400

RESULT 12
Q8N4T0
ID Q8N4T0 PRELIMINARY; PRT; 437 AA.
AC Q8N4T0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to carboxypeptidase A6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033684; AAK33684.1; -.
DR Genew; HGNC:17245; CPA6.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep_M14; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept. 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase.
SQ SEQUENCE 437 AA; 51024 MW; 9970680D429A0A98 CRC64;

Alignment Scores:
Pred. No.: 1.24e-52 Length: 437
Score: 611.50 Matches: 142
Percent Similarity: 49.76% Conservative: 67
Best Local Similarity: 33.81% Mismatches: 140
Query Match: 31.65% Indels: 72
DB: 4 Gaps: 11

US-09-980-881a-1_COPY_18_1097 (1-1080) x Q8N4T0 (1-437)
Qy 1 ATGAAGCTTTC-----AGCTTTCAGTCTTGTACCCCAATTGTTCTTCTGTGAGCAG 54
Db 15 LeuProLeuCysTrpLeuPheLeuLysIleLeuGlnPro-----GlyHisSer 30
Qy 55 CATGCTTC-----GGTTCACAGTGGCCCAAGTCTTAGCTGCTTCTCTAGAACCTCT 108
Db 31 HisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIleProLysThrGlu 50
Qy 109 AGGCAAGTCAAGTTCTACAGAATCTTACTACACATATGATGTTGTTCTCTGCGACCG 168
Db 51 GluGluAlaTyrAlaLeuLysGlyIleSerTyrGlnLeuLysValAspLeuTrpGlnPro 70
Qy 169 GTAACAGCTGACCTTATTGTGAAG-----AAAAACAAGTCCATTTTTTTTGTAAATGCA 222
Db 71 SerSerIleSerTyrValSerGluGlyThrValThrAspValHisIleProGlnAsnGly 90
Qy 223 TCTGATGCGACATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGTCAGTGCTTG 282
Db 91 SerArgAla-----LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyrLysValLeu 108
Qy 283 CTGCGCAGCTGGGAAGATCTTATTCAACAGCAGATTTCCACAGCACAGTCAGCCCCGA 342
Db 109 IleGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGlnArgAsnArg 128
Qy 343 GCCTCCGCATCG-----TACTATGAACAGATGATCACTCACTAAATGAAATCTATTCTTGG 396
```

```
Db 129 ArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIleGlnAsnTrp 148
Qy 397 ATAGAATTTTATAACTGAGAGGCATCTGATATCTTACAAAAATCCACATTGGATCTCA 456
Db 149 MetHisIleLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIleGlyArgSer 168
Qy 457 TTTGAGAGATACCCACTCTATGTTTAAAGTTTCTGGAAGAAAGAACAAACAGCCAAAT 516
Db 169 TyrGluGlyArgCysLeuPheIleLeuLysLeu---GlyArgArgSerArgLeuLysArg 187
Qy 517 GCCATATGGAATTCAGTGTGAATCCATGCCAGAGAATGGATCTCTCTGCTTCTGCTTG 576
Db 188 AlavalTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAlaPheCysGln 207
Qy 577 TGGTTTCATA----- 585
Db 208 TrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArgLysMetLeu 227
Qy 585 ----- 585
Db 228 AsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHisPheSerTrp 247
Qy 586 GGCATAATCGAATGTGGAGAAAGAACCGTTCTTCTATCGAACATCATTCATCGACGA 645
Db 248 ThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArgCysArgGly 267
Qy 646 ACAGACCTGAATAGCAACTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 705
Db 268 ValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAlaSerMetHis 286
Qy 706 TCATGCTCGGAAACCTACTGTGGACTTTTATCTGAGTCAGAACCAAGTGAAGGCAAGTG 765
Db 287 ProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluValLysAlaVal 306
Qy 766 GCTAGTTTCTTGAGAGAATAATCAACAGATTAAGCATACATACAGCATGTCATTCATAC 825
Db 307 AlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPheHisAlaTyr 326
Qy 826 TCCAGCATATAGTGTTCATATCTTATACAGTAAGTAAAGCAAGCAAGCAAGCAAGCA 885
Db 327 AlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsnPheArgCys 346
Qy 886 CTGCTCTCTAGTACGAGTGAAGCAGTTCGTCTATTGACAAAACTAGTAAAAATACCAAG 945
Db 347 ValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyrGlyValArg 365
Qy 946 TATACATAGCCCATGCTCAGAAACCTTATACCTAGTCTCTGAGTGGGAGCATGTTGG 1005
Db 366 TyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSerMetAspTrp 385
Qy 1006 ATCTATGATTTGGGCATCAATATTCGTTTAC----- 1037
Db 386 AlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThrGlyTyrPhe 405
Qy 1038 -----ATCAAAACCCACCTGTAGAGAAGCTTTTTCGCGCTGTC 1073
Db 406 GlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMetLeuAlaVal 425

RESULT 13
Q8AVS2
ID Q8AVS2 PRELIMINARY; PRT; 434 AA.
AC Q8AVS2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to carboxypeptidase A6.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
```

RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (DSC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC041315; AAH41315.1; -;
KW Carboxypeptidase.
SQ SEQUENCE 434 AA; 50581 MW; B42BD77843910591 CRC64;

Alignment Scores:
Pred. No.: 1.75e-52 Length: 434
Score: 610.00 Matches: 145
Percent Similarity: 49.41% Conservativity: 63
Best Local Similarity: 34.44% Mismatches: 134
Query Match: 31.57% Indels: 80
DB: 13 Gaps: 11

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8AVS2 (1-434)

QY 22 GTCTTGTACCCATGTTCTCTCTGTGAGCAGCATGTTCTTGGGTTCCAGAGTGGCCAA 81
DB 11 ValLeuLeuPheLeuLeuSerCys-----AlaLeuSerAlaGlyGln 25

QY 82 -----GTTCTAGCTGCTCTCTCTAGAAC 105
DB 26 CysHisLeuTyrAsnAsnArgTyrSerTrpAspLysValIleArgValPheProLysAsn 45

QY 106 TCTAGGCAAGTTCAAGTTCTACAGAACTTACTACACATATGAGATTGTTCTCTGGCAG 165
DB 46 GluArgGluAlaAsnGluLeuLysAspMetTyrGlnGlnLeuGlnValAspLeuTrpGln 65

QY 166 CCGGTAAACAGCTGACCTTATTGTGAAGAAA-----AAACAAGTCCATTTTGTGTAAT 219
DB 66 ProSerSerIleSerHisIleGlyLysAspThrValThrAspValHisThrSerGlyAsn 85

QY 220 GCATCTGATGTCGAATGTGAAGCCCATTTAAATGTGAGCGGAATTCATGTCAGTGC 279
DB 86 SerSerGluVal-----LeuLeuThrTyrLeuThrLysAlaLysIleGlnHisLysIle 103

QY 280 TTGCTGGCAGACGTGGAGATCTTATTCAA---CAGCAGATTCCACGACACAGTCAGC 336
DB 104 LeuValAsnAsnValGlnSerMetLeuGluAlaGlnGlnAlaPheArgProArgArgLys 123

QY 337 CCCCAGGCTCCGATCTGATC---TATGAACAGTATCACTCACTAAATGAATCTATCT 393
DB 124 ArgArgSerLeuSerArgTyrAsnTyrAsnGluTyrHisProLeuHisGluIleGluSer 143

QY 394 TGGATAGAAATTAATCTAGAGGATCTCTGATATGCTTACAAAATCCATGATGCC 453
DB 144 TrpMetPheTyrMetAsnLysThrHisHisAspLeuValSerLeuPheThrIleGlyLys 163

QY 454 TCATTTGAGAGTACCCTCTATGTTTAAAGTTTCTGGAAGAAAGAACAAACAGCCAAA 513
DB 164 SerTyrGluGlyArgSerLeuTyrValLeuLysLeuGlyLysAspThrLysSerTyrLys 183

QY 514 AATGCTATGATGATGCTGGAATCCATCCAGAGATGATGCTCTCTCTGCTTTCTGC 573
DB 184 LysAlaIleTrpIleAspCysGlyMetHisAlaArgGluTrpIleGlyProAlaPheCys 203

QY 574 TTGTGGTTCAVA----- 585

DB 204 GlnTrpPheValLysGluAlaIleAsnSerTyrAsnThrAspProAlaMetLysLysIle 223

QY 586 -----GGCCAT----- 591

DB 224 LeuAsnLeuLeuTyrIleTyrValMetProValPheAsnValAspGlyTyrHisTyrSer 243

QY 592 -----AATCGAATGTGAGAAAGAACCGTTCTTCTATGCGAACATCATTTGATC 642

DB 244 TrpHisSerAspArgPheTrpArgLysThrArgSerLysAsnThrArgTyrGlnCysTyr 263

QY 643 GGAACAGACCTGGAATAGCACTTTGTCTCCAAACACTGGTGTGAGGAAGTGTGATCCAGT 702

DB 264 GlyValAspAlaAsnArgAsnTrp---LysValHisTrpSerAspGluGlyAlaSerLeu 282

QY 703 TCCTCATGCTCGGAAACCTACTGTGACTTTTCTCTGAGTCAGAACCAAGAGTGAAGCCA 762
DB 283 AsnProCysAspAsnThrTyrCysGlyProTyrAlaGluSerGluProGluValLysAla 302

QY 763 GTGGTAGTTTCTTGAGAGAAATCAACACAGATTAAAGCATACATCATGATGATGCA 822
DB 303 ValAlaGlnPheLeuTyrLysGlnArgLysHisValArgAlaTyrMetSerPheHisAla 322

QY 823 TACTCCAGCATATAGTTTCCATATTCCTATACAGAGTAAAGCAAGAACACCATGAG 882
DB 323 TyrAlaGlnMetLeuLeuTyrProTyrSerTyrGlnTyrGlyAlaIleProAsnPheGly 342

QY 883 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTCTATTGACAAACTAGTAAAAATACC 942
DB 343 CysValGluSerAlaAlaHisAsnAlaValLeuAlaIle---ArgSerAlaTyrGlyIle 361

QY 943 AGGTATACATGATGCCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCAT 1002
DB 362 ArgTyrArgHisGlyProAlaSerSerThrLeuTyrLeuThrSerGlySerSerMetAsp 381

QY 1003 TGGATCTATGATTTGGGCATCAATATTCTGTTTAC----- 1037
DB 382 TrpAlaTyrAsnAsnGlyIleProTyrSerTyrAlaPheGluLeuArgAspThrGlyTyr 401

QY 1038 -----ATCAAAACCCACCTGTAGAGAGCTTTTGGCGCT 1070
DB 402 TyrGlyPheLeuLeuProGluGlyLeuIleLysProThrCysValGluThrMetLeuAla 421

QY 1071 GTC 1073
DB 422 Val 422

RESULT 14
Q8UUK1 PRELIMINARY; PRT; 419 AA.
ID Q8UUK1
AC Q8UUK1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Preprocarboxypeptidase A precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=domesticus; TISSUE=Pancreas;
RA Hasegawa S., Honda K., Hikami Y.;
RT "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A precursor."
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; X64539; CAA45837.1; -;
DR MEROPS; M14.001; -;
DR InterPro; IPR003146; Propep M14.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF02046; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Signal; Carboxypeptidase.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 113 419 CARBOXYPEPTIDASE A.
SQ SEQUENCE 419 AA; 47006 MW; 299C334274BC9C CRC64;

Alignment Scores:
Pred. No.: 1.45e-49 Length: 419
Score: 581.00 Matches: 133
Percent Similarity: 52.52% Conservativity: 65
Best Local Similarity: 35.28% Mismatches: 107

Query Match:	30.07%	Indels:	72
DB:	13	Gaps:	9
US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8UUK1 (1-419)			
QY	67	TTCCAGAGTGGCCGAAGTCTAGCTGCTCTTCTAGAACTCTAGGCAAGTTCAAGTTCTTA	126
Db	21	PheValGlyHisGlnValLeuAlegValProSerSerAspAlaGlnGlnVal	40
QY	127	CAGAACTCTTACTACAAACATATGAGATTGTTCTGCGACCGGTAAACAGCTGACCTATT	186
Db	41	GlnGluLeuGlnGlu	45
QY	187	GTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAAT	237
Db	46	LeuGluHisLeuGlnLeuAspPheTrpLeuSerProArgGlyLeuGlyAsnProValAsp	65
QY	238	-----GTGAAGGCCATTAAATGTGAGCGAATT	267
Db	66	IleArgValProPheProSerLeuGlnProValLysAlaHisLeuGluAlaAsnGlyVal	85
QY	268	CCATGTCAGTGTCTGTCGCGACAGCTGGAAGATCTTATT	315
Db	86	ProTyrSerIleMetIleGluAspValGlnAlaLeuValAspLeuGlnMetGlnMet	105
QY	316	ATTTCACACGACAGTCAGCCCGAGCCTCCGCATCGTAC	372
Db	106	LeuArgArgArgGlyPheValProLeuSerThrSerThrPheAspTyrThrSerTyrHis	125
QY	373	TCACTAAATGAAATCTATCTTGGTAGAATTTATACTGAGAGGCATCTCATATGCTT	432
Db	126	ThrLeuGluGluIleTyrAlaPheMetAspLeuLeuValAlaGluAsnProAsnLeuVal	145
QY	433	ACAAAATCCATTTGGATCTCAFTTCGAGAAGTACCCACTCTATGTTTAAAGTTCT	492
Db	146	SerLysLeuGluIleGlyArgThrThrGluAsnArgProIleTyrValLeuLysPheSer	165
QY	493	GGAAAAGAACAAACAGCCCAAAATGCCATATGATGTGATGTGGAATCCATGCCAGAA	552
Db	166	--LysGlyGlyThrAsnArgProAlaIleTrpIleAspThrGlyIleHisSerArgGlu	184
QY	553	TGATCTCTCTGCTTCTGCTGTGCTTCTATA	585
Db	185	TrpValThrGlnAlaSerGlyIleTrpPheAlaLysLysIleValGlnGlnAspGlu	204
QY	585	-----	585
Db	205	GlyLeuAlaAsnIleLeuAspGlnMetAspIlePheLeuGluIleValThrAsnProAsp	224
QY	586	-----GGCCATAATCGAATGTGGAGAAAGAACCGTTTCTTCTATGCG	627
Db	225	GlyPheAlaPheThrHisThrSerAsnArgMetTrpArgLysThrArgSerLysArgSer	244
QY	628	AACAATCATTCATCGGAACAGACCTGTAATAGCAACTTTGTCTCCAAACACTGGTGTGAG	687
Db	245	GlySerLeuCysValGlyValAspProAsnArgAsnTrp	263
QY	688	GAAGTGCATCCAGTTCCTCATGCTCGAAGACCTACTGTGAGATTATCCCTGAGTCAGAA	747
Db	264	SerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyProTyrAlaAsnSerGlu	283
QY	748	CCGAAGTGAAGGAGTGGCTAGTTTCTTGAGAAAGAAATATCAACAGATTAAAGCATAC	807
Db	284	ProGluValLysAlaIleValAspPheValLysSerHisLysAsn	302
QY	808	ATCAGCATGCATTCATCTCCAGCATATAGTGTTCATATTCCTATACACGAAGTAA	867
Db	303	IleSerIleHisSerTyrSerGlnLeuLeuTyrProTyrGlyTyrThrThrThra	322
QY	868	AGCAAAGACCATGAGGAAGTCTCTCTAGTACGAGTGAAGCGATTCGTCTATTACAA	927
Db	323	ValProAspLysGluLeuHisGlnValAlaLysGluAlaValAlaAlaLeu	341

Qy	928	ACTAGTAA	AAATACACAGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCT	98
Db		:::	::: ::: :::	
Db	342	SerLeu	TyrGlyThrAsnTyrSerGlySerIleIleThrThrIleTyrGlnAlaSer	361
Qy	988	GGAGGTGGG	ACGAGTTCGATCTATCATTTGGGCATCAAAATTCGTTTACA	1038
Db			::: ::: ::: :::	
Db	362	GlyGlyThrIle	AspTrpThrTyrAsnGlnGlyIleTyrSerPheThr	378
RESULT 15				
Q81VL8				
ID	Q81VL8	PRELIMINARY;	PRT;	374 AA.
AC	Q81VL8;			
DT	01-MAR-2003	(TrEMBLrel. 23, Created)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)		
DE	Zn-carboxypeptidase.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RA	Obaya A.J., Lopez-Otin C.;			
RT	"A new Zn-carboxypeptidase highly expressed in ovary.";			
RL	Submitted (DBC-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AJ422118; CAD19478.1; -			
KW	Carboxypeptidase.			
SQ	SEQUENCE	374 AA; 42529 MW; 404C373B8841AAD2 CRC64;		
Alignment Scores:				
Pred. No.:	6.49e-48	Length:	374	
Score:	564.50	Matches:	115	
Percent Similarity:	53.70%	Conservative:	52	
Best Local Similarity:	36.98%	Mismatches:	87	
Query Match:	29.22%	Indels:	58	
DB:	4	Gaps:	6	
US-09-980-881A-1_COPY_18_1097 (1-1080) x Q81VL8 (1-374)				
Qy	307	CAACGACGATTTCC	AACGACACAGTCAGCCGCCGAGCTCCGCATCGTAC---	TATGNA 363
Db		::: ::: ::: ::: :::		
Db	29	ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn	48	
Qy	364	CAGTACTCACTCACTAAATGAATCTATCTCTGGATAGAATTTATAACTGAGAGGACCTCT	423	
Db				
Db	49	IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys	68	
Qy	424	GATATGCTTACAAAATCCACATTTGGATCTCATTTGAGAGTACCCACTCTATGTTT	483	
Db		::: ::: ::: ::: :::		
Db	69	GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProMetTyrTyrLeu	88	
Qy	484	AAGTTTCTGAAAAGAACAAACAGCCAAAATGCCATATGATTCGATTCGGAATCCAT	543	
Db				
Db	89	LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis	108	
Qy	544	GCCAGAGATGGATCTCTCCGCTTTCTGCTTGTGGTTCAATA-----	585	
Db				
Db	109	AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn	128	
Qy	585	-----	-----	585
Db	129	HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro	148	
Qy	586	-----GGCCAT-----	-----AATCGAATGTGGAGAAGAAC	612
Db				
Db	149	ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer	168	
Qy	613	CGTTCTTTCTATGCGAACAAATCATTGCTCGGACAGACCTGGAATAGCAACTCTCTCC	672	
Db				
Db	169	ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---	Asn 187	
Qy	673	AAACACTGGTGTGAGGAGGTGCATCCATGTTCTCTCATCTCGGAAACCTACTGTGGACTT	732	


```
Db 188 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 207
Qy 733 TATCCTGAGTCAGAACAGAGTGAAGCGAGTGTCTTCTTGAGAAGAAATATCAAC 792
Db 208 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 227
Qy 793 CAGATTAAAGCATACATCAGCATGCATTTCATCTCCAGCATATAGTGTTCCTCATATCC 852
Db 228 AspileLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 247
Qy 853 TATACAGAGTAAGCAAGACCATGAGGAAGTGTCTCTAGTACGAGTGAAGCAGTT 912
Db 248 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 267
Qy 913 CGTCTATTGACAAACTAGTAAATAACACAGGTATACATGCGCCATGGCTCAGAAACC 972
Db 268 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspile 286
Qy 973 TTATACCTAGTCTCTGGAGTGGGAGCATTTGGATCTATGATTTGGGCATCAATATTCG 1032
Db 287 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspileGlyIleProPheSer 306
Qy 1033 TTTACATC----- 1040
Db 307 TyrThrPheGluLeuArgAspSerGlyThrTyrGlyPheValLeuProGluAlaGlnIle 326
Qy 1041 AAACCCACCTGTAGAGAGCTTTTCCCGCTGTC 1073
Db 327 GlnProThrCysGluGluThrMetGluAlaVal 337
```

Search completed: January 1, 2004, 18:58:53
Job time : 74.6425 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:58:24 ; Search time 4065.16 Seconds
(without alignments)
10868.540 Million cell updates/sec

Title: US-09-980-881a-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagcttgagccttg.....ttttgcgcgtgtctctaaaa 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1064	98.5	1546	9	AB011969	AB011969 Homo sapi
2	1025.2	94.9	1625	6	AR232183	AR232183 Sequence
3	909	84.2	1272	6	AR086324	AR086324 Sequence
4	909	84.2	1272	6	BD106653	BD106653 Method of
5	909	84.2	1728	6	AX409750	AX409750 Sequence
6	909	84.2	1728	6	AX706330	AX706330 Sequence
7	909	84.2	1728	9	HUMPCPBX	M75106 Human prepr
8	909	84.2	1749	6	I16100	I16100 Sequence 2
9	909	84.2	1749	6	I33526	I33526 Sequence 2
10	907.4	84.0	1272	9	BT006936	BT006936 Homo sapi
11	907.4	84.0	1272	12	BT007990	BT007990 Synthetic
12	907.4	84.0	1715	9	BC007057	BC007057 Homo sapi
13	643.4	59.6	1421	10	AB021968	AB021968 Mus muscu
14	643.4	59.6	1490	10	AF164524	AF164524 Mus muscu
15	642.4	59.5	1425	10	AB042598	AB042598 Rattus no
16	638.6	59.1	1430	10	AF186188	AF186188 Mus muscu
17	204	18.9	55827	6	AR232184	AR232184 Sequence
18	204	18.9	58097	9	AL157758	AL157758 Human DNA
19	156	14.4	251806	2	AC096305	AC096305 Rattus no
20	156	14.4	258265	2	AC120472	AC120472 Rattus no
21	129.6	12.0	431	6	AX400792	AX400792 Sequence
22	128	11.9	180680	9	AL137141	AL137141 Human DNA
23	124	11.5	927	6	AR072911	AR072911 Sequence
24	123.8	11.5	1327	5	AF190274	AF190274 Bothrops
25	123.4	11.4	921	6	I67699	I67699 Sequence 3
26	123.4	11.4	1215	6	I67698	I67698 Sequence 1
27	123.4	11.4	1251	4	SSC133775	AJ133775 Sus scrof
28	123.2	11.4	451	11	G53326	G53326 SHGC-82320
29	120.8	11.2	1251	4	DOGZAP47	D78348 Canis famli
30	109	10.1	1623	6	AR7530	AR7530 Sequence 11
31	107.6	10.0	1622	9	HUMCARMC	M27717 Human mast
32	107.6	10.0	1622	11	G28614	G28614 human STS S
33	107.6	10.0	1674	9	BC012613	BC012613 Homo sapi
34	107.4	9.9	999	6	A51908	A51908 Sequence 72
35	107.4	9.9	999	6	AR085871	AR085871 Sequence
36	107.4	9.9	1053	6	A51904	A51904 Sequence 68
37	107.4	9.9	1053	6	AR085868	AR085868 Sequence
38	107.4	9.9	1254	12	BT008111	BT008111 Synthetic
39	107.4	9.9	1263	6	A51896	A51896 Sequence 60
40	107.4	9.9	1263	6	AR085861	AR085861 Sequence
41	107.4	9.9	1284	6	A51913	A51913 Sequence 77
42	107.4	9.9	1284	6	AR085875	AR085875 Sequence
43	107.4	9.9	1310	9	HSA224866	AJ224866 Homo sapi
44	107.4	9.9	1454	9	BC015338	BC015338 Homo sapi
45	104.2	9.6	1053	6	AR300428	AR300428 Sequence

ALIGNMENTS

RESULT 1	AB011969	1546 bp	mRNA	linear	PRI 02-FEB-2000
LOCUS	AB011969				
DEFINITION	Homo sapiens mRNA for carboxypeptidase B-like protein, complete cds.				
ACCESSION	AB011969				
VERSION	AB011969.1	GI:6855463			
KEYWORDS	carboxypeptidase B-like protein.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (sites)				
AUTHORS	Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.				
TITLE	Isolation, molecular cloning, and partial characterization of a				

Qy	61	TTCCGGTTTCAGAGTGGCCAAAGTTCTTAGCTGCTCTTCTAGTAAACCTCTAGGCAAGTTCAA	120
Db	77	TTCCGGTTTCAGAGTGGCCAAAGTTCTTAGCTGCTCTTCTAGTAAACCTCTAGGCAAGTTCAA	136
Qy	121	GTTCTCAGAGATCTTTACTACAACATATGAGATGCTTCTCTGGCAGCGGTAAACAGCTGAC	180
Db	137	GTTCTCAGAGATCTTTACTACAACATATGAGATGTTTCTCTGGCAGCGGTAAACAGCTGAC	196
Qy	181	CTTATTGTGAAGAAAAAACAAGTCCATTTTTTGTCTAAATGCATCTGATTCGACAAATGTG	240
Db	197	CTTATTGTGAAGAAAAAACAAGTCCATTTTTTGTCTAAATGCATCTGATTCGACAAATGTG	256
Qy	241	AAAGCCCATTTAAATGTGAGCGGAATTCATCGAGTGTCTTGCTGGCAGAGCTGGAAGAT	300
Db	257	AAAGCCCATTTAAATGTGAGCGGAATTCATCGAGTGTCTTGCTGGCAGAGTGTGGAAGAT	316
Qy	301	CTTATTCAACAGCAGATTTTCAACGACACACAGTCAAGCCCCGAGCCTCCGATCCGTACTAT	360
Db	317	CTTATTCAACAGCAGATTTTCAACGACACACAGTCAAGCCCCGAGCCTCCGATCCGTACTAT	376
Qy	361	GACAGATATCACTCACTAAATGAAATCTATTTCTTGGATAGAAATTTATACTGAGAGGCAT	420
Db	377	GACAGATATCACTCACTAAATGAAATCTATTTCTTGGATAGAAATTTATACTGAGAGGCAT	436
Qy	421	CCTGATATGCTTACAAAATCCACATTTGGATCCCTCATTTGAGAGTAGTACCACCTCTATGTT	480
Db	437	CCTGATATGCTTACAAAATCCACATTTGGATGCTCTCATTTGAGAGTAGTACCACCTCTATGTT	496
Qy	481	TTAAAGGTTTCTGAAAAGAAACAAACAGCCAAAATGCCATATGGAATGACTGTGGAATC	540
Db	497	TTAAAGGTTTCTGAAAAGAAACAAACAGCCAAAATGCCATATGGAATGACTGTGGAATC	556
Qy	541	CATGCCAGAGAAATGGATCTCTCTGCTTTCTGCTTGTGGTTTCATPAGGCCATAATCGAATG	600
Db	557	CATGCCAGAGAAATGGATCTCTCTGCTTTCTGCTTGTGGTTTCATPAGGCCATAATCGAATG	616
Qy	601	TGGAGAAAGAACCGTTCTTTCTATGCGAAACAATCATTTGCATCGGAAACAGACCTGAAATAGC	660
Db	617	TGGAGAAAGAACCGTTCTTTCTATGCGAAACAATCATTTGCATCGGAAACAGACCTGGAATAGG	676
Qy	661	AACTTTGTCTCAAAACACTGGTGTGAGGAAGGTGCATCCAGTTTCCTCATGCTCGGAAACC	720
Db	677	AACTTTGTCTCAAAACACTGGTGTGAGGAAGGTGCATCCAGTTTCCTCATGCTCGGAAACC	736
Qy	721	TACTGTGGACTTTATCTGTAGTCAAGACCAAGAGTGAAGGCAGTGGCTAGTTTCTTGAGA	780
Db	737	TACTGTGGACTTTATCTGTAGTCAAGACCAAGAGTGAAGGCAGTGGCTAGTTTCTTGAGA	796
Qy	781	AGAAATATCAACAGATTTAAAGCATACATCAGCATGCATTCATACTCCAGCATATAGTG	840
Db	797	AGAAATATCAACAGATTTAAAGCATACATCAGCATGCATTCATACTCCAGCATATAGTG	856
Qy	841	TTTCCATATTCCTTATACAGGAATAAAGCAAGAACCATGAGGAACCTGTCTCTAGTAGGCC	900
Db	857	TTTCCATATTCCTTATACAGGAATAAAGCAAGAACCATGAGGAACCTGTCTCTAGTAGGCC	916
Qy	901	AGTGAAGCAGTTCGTGCTATTGACAAACCTAGTAAAATAACAGGTATACACATGGCCAT	960
Db	917	AGTGAAGCAGTTCGTGCTATTGAGAAAATTTAGTAAAAATAACAGGTATACACATGGCCAT	976
Qy	961	GGCTCAGAAAACCTTATACCTAGTCTCTGGAGGTGGGACGATTTGGATCTATGATTTGGGC	1020
Db	977	GGCTCAGAAAACCTTATACCTAGTCTCTGGAGGTGGGACGATTTGGATCTATGATTTGGGC	1036
Qy	1021	ATCAAAATATTCGTTTACA	1038
Db	1037	ATCAAAATATTCGTTTACA	1054

RESULT 3	DEFINITION
AR086324	
LOCUS	

AR086324 1272 bp DNA linear PAT 07-SEP-2000
Sequence 1 from patent US 5985562.

ACCESSION	AR086324
VERSION	AR086324.1 GI:10013090
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1. (bases 1 to 1272)
TITLE	Morser, M.John. and Nagashima, M. Method of detecting thrombotic disease risk associated with plasma carboxypeptidase B polymorphisms
JOURNAL	Patent: US 5985562-A 1 16-NOV-1999;
FEATURES	Location/Qualifiers 1..1272
source	/organism="unknown"
BASE COUNT	375 a 269 c 271 g 357 t
ORIGIN	
Query Match	84.2%; Score 909; DB 6; Length 1272;
Best Local Similarity	89.9%; Pred. No. 1.9e-246;
Matches 1033; Conservative	0; Mismatches 5; Indels 111; Gaps 1;
Qy	1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGATACCATTGTTCTCTTCTGTGAGCAGCATGTC 60
Dd	1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGATACCATTGTTCTCTTCTGTGAGCAGCATGTC 60
Qy	61 TTCGGTTCCAGAGTGGCCAAAGTTCCTAGCTGCTCTCTTAGAACAACCTCTAGGCCAAGTTCAA 120
Dd	61 TTCGGTTTCAGAGTGGCCAAAGTTCCTAGCTGCTCTCTTAGAACAACCTCTAGGCCAAGTTCAA 120
Qy	121 GTTCTACAGAATCTTACTCAACATATGAGATGTTCTCTGGCAGCGCGGTAAACAGCTGAC 180
Dd	121 GTTCTACAGAATCTTACTCAACATATGAGATGTTCTCTGGCAGCGCGGTAAACAGCTGAC 180
Qy	181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAATGCATCTGATGTCGACAATGTG 240
Dd	181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAATGCATCTGATGTCGACAATGTG 240
Qy	241 AAAGCCCATTTAAATGTGAGCGGAATTCNATGAGTGTCTTGCTGGCAGACGCTGGGAAGAT 300
Dd	241 AAAGCCCATTTAAATGTGAGCGGAATTTCCAATGAGTGTCTTGCTGGCAGACGCTGGGAAGAT 300
Qy	301 CTTATTCAACAGCAGATTTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCCTACTAT 360
Dd	301 CTTATTCAACAGCAGATTTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCCTACTAT 360
Qy	361 GAACAGTATCACCTCACTAAATGAAATCTATTCTTGATAGAAATTTATAACTGAGAGGCAT 420
Dd	361 GAACAGTATCACCTCACTAAATGAAATCTATTCTTGATAGAAATTTATAACTGAGAGGCAT 420
Qy	421 CCTGATATGCTTACAAAATCCACATTTGNATCTCATTTTGAGAAGTACCCACCTATGTT 480
Dd	421 CCTGATATGCTTACAAAATCCACATTTGGATCTCTATTGAGAAGTACCCACCTCTATGTT 480
Qy	481 TTAAGGTTTCTCGAAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTCTGGAATC 540
Dd	481 TTAAGGTTTCTCGAAAAAGNAACAAACAGCCAAAAATGCCATATGGATTGACTGTGNATC 540
Qy	541 CATGCCAGAAATGGATCTCTCTGCTTTTCTGCTTTGTTGTTTCATAGGCCAT----- 591
Dd	541 CATGCCAGAAATGGATCTCTCTGCTTTTCTGCTTTTCTGCTTTTCTAGGCCATATAAATCTCAA 600
Qy	592 ----- 591
Dd	601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTTGGAATTTCTATGTATG 660
Qy	592 -----AATCGAATGTGGAGAAG 609
Dd	661 CCGGTGGTTAACTGGACGGTTATGACTACTCATGGAAAAAAGAAATCGAATGTGGAGAAAG 720
Qy	610 AACCGTTCTTTCTATGCGAACAAATCAATTCGATCGGAACAGACTCGAATACCAACTTTGTC 669
Dd	721 AACCGTTCTTTCTATGCAACATCAATTCGATCGGAACAGACTCGAATAGGAACCTTTGCT 780

```
QY 670 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCATGCTCGGAAACCTACTGTGGA 729
DB 781 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCATGCTCGGAAACCTACTGTGGA 840
QY 730 CTTTATCTGAGTCAGAACCCAGAAAGTGAAGCAGTGGCTAGTTCTTTGAGAAGAAATATC 789
DB 841 CTTTATCTGAGTCAGAACCCAGAAAGTGAAGCAGTGGCTAGTTCTTTGAGAAGAAATATC 900
QY 790 AACAGATTAAGCATACATCAGCATGAGTATCATCTCCAGCATATAGTTTCCATAT 849
DB 901 AACAGATTAAGCATACATCAGCATGAGTATCATCTCCAGCATATAGTTTCCATAT 960
QY 850 TCCTATACAGAACTAAAGCAACAGACATGAGGAACTGTCTCTAGTAGCAGTGAAGCA 909
DB 961 TCCTATACAGAACTAAAGCAACAGACATGAGGAACTGTCTCTAGTAGCAGTGAAGCA 1020
QY 910 GTTCGTGCTATTGACAAACCTAGTAAATAACCAAGGTATACATAGGCCATGGCTCAGAA 969
DB 1021 GTTCGTGCTATTGAGAAACCTAGTAAATAACCAAGGTATACATAGGCCATGGCTCAGAA 1080
QY 970 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCACTCAATAT 1029
DB 1081 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCACTCAATAT 1140
QY 1030 TCGTTTACA 1038
DB 1141 TCGTTTACA 1149

RESULT 4
BD106653 LOCUS 1272 bp DNA linear PAT 18-SEP-2002
DEFINITION Method of detecting thrombotic disease risk.
ACCESSION BD106653
VERSION BD106653.1 GI:23201471
KEYWORDS JP 2002502253-A/1.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1272)
AUTHORS Morser,M.J. and Nagashima,M.
TITLE Method of detecting thrombotic disease risk
JOURNAL Patent: JP 2002502253-A 1 22-JAN-2002;
SCHERING AG
COMMENT PN JP 2002502253-A/1
PD 22-JUN-2002
PF 02-JUN-1998 JP 1999501465
PR 03-JUN-1997 US 08/869057
PI MICHAEL JOHN MORSER,MARIKO NAGASHIMA
PC C1201/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..1272
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN

Query Match 84.2%; Score 909; DB 6; Length 1272;
Best Local Similarity 89.9%; Pred. No. 1.9e-246;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

QY 1 ATGAAGCTTTGACGCTTGCAGTCCCTGTACCCATTGCTCTCTCTGAGCAGCATGTC 60
DB 1 ATGAAGCTTTGACGCTTGCAGTCCCTGTACCCATTGCTCTCTCTGAGCAGCATGTC 60
QY 61 TTCGCGTTCCAGAGTGCCCAAGTTCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
DB 61 TTCGCGTTCCAGAGTGCCCAAGTTCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
```

```
QY 121 GTTCTCAGAACTCTTACTACAAATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
DB 121 GTTCTCAGAACTCTTACTACAAATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTCTGATGTCGACAAATGTG 240
DB 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTCTGATGTCGACAAATGTG 240
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTCTGGCAGACGTGGGAAGAT 300
DB 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTCTGGCAGACGTGGGAAGAT 300
QY 301 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCAGCCCCCGAGCCCTCCGATCGTACTAT 360
DB 301 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCAGCCCCCGAGCCCTCCGATCGTACTAT 360
QY 361 GAACAGTATCAGTCACTCACTAAATGAAATCTATTCTTGGATGAAATTTATACTGAGAGGCAT 420
DB 361 GAACAGTATCAGTCACTCACTAAATGAAATCTATTCTTGGATGAAATTTATACTGAGAGGCAT 420
QY 421 CTTGATATGCTTACAAAATTCACATTTGGATCTCTATTTGAGAAGTACCCACTCTATGTT 480
DB 421 CTTGATATGCTTACAAAATTCACATTTGGATCTCTATTTGAGAAGTACCCACTCTATGTT 480
QY 481 TTAAGGTTTCTGGAAGAAACAAACAGCCAAATGCCCATATGGATTTGACTGTGGAATC 540
DB 481 TTAAGGTTTCTGGAAGAAACAAACAGCCAAATGCCCATATGGATTTGACTGTGGAATC 540
QY 541 CATGCCAGAGATGGATCTCTCTGCTTCTGCTGTGTTCTATAGGCCAT----- 591
DB 541 CATGCCAGAGATGGATCTCTCTGCTTCTGCTGTGTTCTATAGGCCATATAACTCAA 600
QY 592 ----- 591
DB 601 TTCTATGGGATAATAGGGCAATATACCAATCTCTCGAGGCTTGTGGATTTCTATGTTATG 660
QY 592 -----NATCGAATGTCGGAAG 609
DB 661 CCGGTGGTTAAATGTGACGGTTATGACTACTCATGGAAGAAATCGAATGTGGAAGAAAG 720
QY 610 AACCGTTCTTTCTATCGAACAATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
DB 721 AACCGTTCTTTCTATCGGAACATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGTC 780
QY 670 TCCAAACACTGTTGTGAGGAAGTGCATCCAGTTCTCTATGCTCGGAAACCTACTGTGGA 729
DB 781 TCCAAACACTGTTGTGAGGAAGTGCATCCAGTTCTCTATGCTCGGAAACCTACTGTGGA 840
QY 730 CTTTATCTCAGTCAGAACCCAGAAAGTGAAGCAGTGGCTAGTTCTTTGAGAAGAAATATC 789
DB 841 CTTTATCTCAGTCAGAACCCAGAAAGTGAAGCAGTGGCTAGTTCTTTGAGAAGAAATATC 900
QY 790 AACAGATTAAGCATACATCAGCATGAGTTCATATCTCCAGCATATAGTTTCCATAT 849
DB 901 AACAGATTAAGCATACATCAGCATGAGTTCATATCTCCAGCATATAGTTTCCATAT 960
QY 850 TCCTATACAGAACTAAAGCAACAGACATGAGGAACTGTCTCTAGTAGCAGTGAAGCA 909
DB 961 TCCTATACAGAACTAAAGCAACAGACATGAGGAACTGTCTCTAGTAGCAGTGAAGCA 1020
QY 910 GTTCGTGCTATTGACAAACCTAGTAAATAACCAAGGTATACATAGGCCATGGCTCAGAA 969
DB 1021 GTTCGTGCTATTGAGAAACCTAGTAAATAACCAAGGTATACATAGGCCATGGCTCAGAA 1080
QY 970 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCACTCAATAT 1029
DB 1081 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCACTCAATAT 1140
QY 1030 TCGTTTACA 1038
DB 1141 TCGTTTACA 1149
```

```
RESULT 5
AX409750
LOCUS AX409750 1728 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2397 from Patent WO0229103.
ACCESSION AX409750
VERSION AX409750.1 GI:21442455
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2397 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source
Location/Qualifiers
1..1728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="EMBL/GenBank Accession No. M75106"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN
Query Match 84.2%; Score 909; DB 6; Length 1728;
Best Local Similarity 89.9%; Pred. No. 1.9e-246;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;
QY 1 ATGAAGCTTTGCAGCTTGCAGTCCCTGTGATCCCATTTCTCTCTCTGTGAGCAGCATGTC 60
DB 20 ATGAAGCTTTGCAGCTTGCAGTCCCTGTGATCCCATTTCTCTCTCTGTGAGCAGCATGTC 79
QY 61 TTCGCGTTCCAGAGTGCCCAAGTTCCTAGCTGCTCTTCTAGAGACCTCTAGGCAAGTTCAA 120
DB 80 TTCTATGGGTAATATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 679
RESULT 6
AX706330
LOCUS AX706330 1728 bp DNA linear PAT 04-APR-2003
DEFINITION Sequence 1 from Patent WO03014391.
ACCESSION AX706330
VERSION AX706330.1 GI:29562754
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Morten,J.E. and Jackson,K.N.
TITLE Method
JOURNAL Patent: WO 03014391-A 1 20-FEB-2003;
AstraZeneca AB (SE)
FEATURES
source
Location/Qualifiers
1..1728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN
Query Match 84.2%; Score 909; DB 6; Length 1728;
Best Local Similarity 89.9%; Pred. No. 1.9e-246;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;
QY 1 ATGAAGCTTTGCAGCTTGCAGTCCCTGTGATCCCATTTCTCTCTCTGTGAGCAGCATGTC 60
DB 20 ATGAAGCTTTGCAGCTTGCAGTCCCTGTGATCCCATTTCTCTCTCTGTGAGCAGCATGTC 79
QY 61 TTCGCGTTCCAGAGTGCCCAAGTTCCTAGCTGCTCTTCTAGAGACCTCTAGGCAAGTTCAA 120
DB 80 TTCTATGGGTAATATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 139
```

```
121 GTTCTACAGAACTTTACTCAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 180
140 GTTCTACAGAACTTTACTCAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 199
181 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGATCGCAATGTG 240
199 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGATCGCAATGTG 259
241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTCTTTGCTGGCAGACGTTGAAGAT 300
260 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTCTTTGCTGGCAGACGTTGAAGAT 319
301 CTTATTCAACAGCAGATTTTCAACGACACAGTACAGCCCCCGAGCCTCCGATCGTACTAT 360
320 CTTATTCAACAGCAGATTTTCAACGACACAGTACAGCCCCCGAGCCTCCGATCGTACTAT 379
361 GAACAGTATCACTCACTAAATGAATCTATTCTTTGGATGAATTTATACTGAGAGCAT 420
380 GAACAGTATCACTCACTAAATGAATCTATTCTTTGGATGAATTTATACTGAGAGCAT 439
421 CTTGATATGCTTACAAAAATCCCATTTGGATCTCTATTTCAGAGTACCCACTCTATGTT 480
440 CTTGATATGCTTACAAAAATCCCATTTGGATCTCTATTTCAGAGTACCCACTCTATGTT 499
481 TTTAAAGGTTTCTGGAAGAAACAAACAGCAGTACGAGTATGATGATGATGATGATGATG 540
500 TTTAAAGGTTTCTGGAAGAAACAAACAGCAGTACGAGTATGATGATGATGATGATGATG 559
541 CATGCCAGAGATGATCTCTCTGCTTTCTGCTTTGTTGTTGTTGTTGTTGTTGTTGTTG 591
560 CATGCCAGAGATGATCTCTCTGCTTTCTGCTTTGTTGTTGTTGTTGTTGTTGTTGTTG 619
592 ----- 591
620 TTTATGGGATATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 679
592 ----- 591
680 CCGTGTGTTAATGTGGACGGTTATGACTACTCATCGAAGAAAGAAATCGAATGTGGAGAA 739
610 AACGGTTCTTTCTATGCGAACAATCATGATGATGATGATGATGATGATGATGATGATGAT 669
740 AACGGTTCTTTCTATGCGAACAATCATGATGATGATGATGATGATGATGATGATGATGAT 799
760 TCCAAACACGTGCTGAGGAGGTGATCCAGTCTCTGCTCGGAAACCTACTGTTGGA 729
800 TCCAAACACGTGCTGAGGAGGTGATCCAGTCTCTGCTCGGAAACCTACTGTTGGA 859
730 CTTTATCTGAGTCAGAACAGAGTGAAGCAGTGTGCTAGTCTTCTGAGAGAAATATC 789
860 CTTTATCTGAGTCAGAACAGAGTGAAGCAGTGTGCTAGTCTTCTGAGAGAAATATC 919
790 AACAGATTAAGCATACATCAGATGATCTATCTCCAGCATATAGTGTGTTCCATAT 849
920 AACAGATTAAGCATACATCAGATGATCTATCTCCAGCATATAGTGTGTTCCATAT 979
850 TCCTATACAGAGTAAAGCAAGCATGAGCACTGCTCTAGTGTGAGCAGTGAAGCA 909
980 TCCTATACAGAGTAAAGCAAGCATGAGCACTGCTCTAGTGTGAGCAGTGAAGCA 1039
910 GTTCGTCTATTTGACAAAATAGTAAATAATACAGGTATATACATGAGCCTGCTCAGAA 969
1040 GTTCGTCTATTTGACAAAATAGTAAATAATACAGGTATATACATGAGCCTGCTCAGAA 1099
970 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGATCTATGATTTGGGCATCAATAT 1029
1100 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGATCTATGATTTGGGCATCAATAT 1159
1030 TCCTTTTACA 1038
1160 TCCTTTTACA 1168
```

```
RESULT 7
LOCUS HUMPCPBX
DEFINITION Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION M75106
VERSION M75106.1 GI:189686
KEYWORDS plasma carboxypeptidase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
J. Biol. Chem. 266 (32), 21833-21838 (1991)
JOURNAL 92042093
MEDLINE 1939207
PUBMED 1939207
COMMENT Original source text: Homo sapiens liver cdna to mRNA.
```

FEATURES

source

1..1728

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="liver"

1..1728

/gene="pcPB"

20..1291

/gene="pcPB"

/note="prepro-plasma carboxypeptidase B"

/codon_start=1

/protein_id="AAA60042.1"

/db_xref="GI:189687"

/translation="MKICSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQLQN

LTTTIEVLWQPVTDLIVKKQVHFVNADVNVKHLNVSGIPSPCVLLADVEDLI

QQISNDTVSPRASAYEYQYHSLNEIYSWIFITERHPDMLTKIHGSSFEKYPYLV

LKVSGEQTKAKNAIWIDCGIHAREWISPAFLWFIHQFYGIIGQYTNLRLLVDFY

VMVNVNVDGYDSWKNRMWRNRSFYANNHCIGTDLNRNFASKHWCERGASSSCSSE

TVGGLYPESEPEKAVASFLRNINQIKAYISMHSYSOHIVFPYKSTRKSKDHESLS

LVASEAVRAIEKTSKNTRYTHGSETLYLAPGGDDWIYDLGIKISFIELRDTGT

GFLLPERYIKPTCREAFAVSKIAHWIRNV"

86..1288

/gene="pcPB"

/product="plasma carboxypeptidase B"

518 a 354 c 338 g 518 t

mat_peptide

BASE COUNT 518 a 354 c 338 g 518 t

ORIGIN

Query Match 84.2%; Score 909; DB 9; Length 1728;

Best Local Similarity 89.9%; Pred. No. 1.9e-246;

Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

1 ATGAAGCTTTGAGCGCTTGCAGTCTTGTACCCATTTGTTCTCTTGTGAGCAGATGTC 60

20 ATGAAGCTTTGAGCGCTTGCAGTCTTGTACCCATTTGTTCTCTTGTGAGCAGATGTC 79

61 TTCGCGTTCAGAGTGCAGCAAGTTCTAGCTGCTTCTTAGACCTCTAGGCAAGTTCAA 120

80 TTCGCGTTCAGAGTGCAGCAAGTTCTAGCTGCTTCTTAGACCTCTAGGCAAGTTCAA 139

121 GTTCTACAGAACTTTACTCAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 180

140 GTTCTACAGAACTTTACTCAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 199

181 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGATCGCAATGTG 240

200 CTTATTGTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGATCGCAATGTG 259

241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTCTTTGCTGGCAGCGGTAACAGAT 300

260 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTCTTTGCTGGCAGCGGTAACAGAT 319

301 CTTATTCAACAGCAGATTTTCAACGACACAGTACAGCCCCCGAGCCTCCGATCGTACTAT 360

QY 850 TCCTATACAGAGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 909
Db 1001 TCCTATACAGAGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 1060
QY 910 GTTCGTGCTATTGACAAAACCTAGTAAATAACACAGGTATACACATGGCCATGGCTCAGAA 969
Db 1061 GTTCGTGCTATTGAGAAAACCTAGTAAATAACACAGGTATACACATGGCCATGGCTCAGAA 1120
QY 970 ACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1029
Db 1121 ACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1180
QY 1030 TCGTTTACA 1038
Db 1181 TCGTTTACA 1189

RESULT 9
133526
LOCUS 133526 1749 bp DNA linear PAT 06-FEB-1997
DEFINITION Sequence 2 from patent US 5593674.
ACCESSION 133526
VERSION 133526.1 GI:1824317
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Plasma carboxypeptidase
JOURNAL Patent: US 5593674-A 2 14-JAN-1997;
FEATURES
source Location/Qualifiers
1.1749
/organism="unknown"
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN

Query Match 84.28; Score 909; DB 6; Length 1749;
Best Local Similarity 89.94; Pred. No. 1.9e-246;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

QY 1 ATGAAGCTTTGCGAGCTTCAGTCTCTGTGATCCCAATGTTCTCTCTGAGCAGCATGTC 60
Db 41 ATGAAGCTTTGCGAGCTTCAGTCTCTGTGATCCCAATGTTCTCTCTGAGCAGCATGTC 100
QY 61 TTCGCTTCCAGAGTGCCCAAGTCTAGTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Db 101 TTCGCTTCCAGAGTGCCCAAGTCTAGTCTCTCTAGAACCTCTAGGCAAGTTCAA 160
QY 121 GTTCTACAGATCTTACTACACATATGAGATGTTCTCTGGCAGCCGGTAAACAGCTGAC 180
Db 161 GTTCTACAGATCTTACTACACATATGAGATGTTCTCTGGCAGCCGGTAAACAGCTGAC 220
QY 181 CTTATTCTGGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 240
Db 221 CTTATTCTGGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 280
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTTGCTGGCAGACGTGGAAGAT 300
Db 281 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTTGCTGGCAGACGTGGAAGAT 340
QY 301 CTTATTCAACAGCAGATTTTCAACGACACATGACGCCCCCGAGCCTCCGATCGTACTAT 360
Db 341 CTTATTCAACAGCAGATTTTCAACGACACATGACGCCCCCGAGCCTCCGATCGTACTAT 400
QY 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 420
Db 401 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 460
QY 421 CCTGATATGCTTACAAAAATCCCAATGGATTCCTCATTTGAGAAGTACCCACTCTATGTT 480
Db 461 CCTGATATGCTTACAAAAATCCCAATGGATTCCTCATTTGAGAAGTACCCACTCTATGTT 520

QY 481 TTAAAGGTTTCTGGAAGAAACAAACAGCCAAATATGCCATATGGAATGACTCTGGAATC 540
Db 521 TTAAAGGTTTCTGGAAGAAACAAACAGCCAAATATGCCATATGGAATGACTCTGGAATC 580
QY 541 CATGCCAGAGAATGGATCTCTCTGCTTTTCTGCTTCTGCTTCTCATAGGCCAT----- 591
Db 581 CATGCCAGAGAATGGATCTCTCTGCTTTTCTGCTTCTGCTTCTCATAGGCCATATAACTCAA 640
QY 592 ----- 591
Db 641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGATTTCTATGTTATG 700
QY 592 -----AATCGAATGTGGAGAAAG 609
Db 701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGGAGAAAG 760
QY 610 AACCGTTCTTTCTATCGGAACAATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGTG 669
Db 761 AACCGTTCTTTCTATCGGAACAATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGTG 820
QY 670 TCCAAACACTGTTGTGAGGAAGGTGATCCAGTTCTCTCATGCTCGGAAACCTTACTGTGGA 729
Db 821 TCCAAACACTGTTGTGAGGAAGGTGATCCAGTTCTCTCATGCTCGGAAACCTTACTGTGGA 880
QY 730 CTTTATCCTGAGTCAGAACCCAGAAAGTGAAGGAGTGGCTAGTTCTTCTGAGAAGAAATATC 789
Db 881 CTTTATCCTGAGTCAGAACCCAGAAAGTGAAGGAGTGGCTAGTTCTTCTGAGAAGAAATATC 940
QY 790 AACAGATTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATAT 849
Db 941 AACAGATTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATAT 1000
QY 850 TCCATATACAGAGTAAAGCAAGCAACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCA 909
Db 1001 TCCATATACAGAGTAAAGCAAGCAACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCA 1060
QY 910 GTTCGTGCTATTGACAAAACCTAGTAAATAACACAGGTATACACATGGCCATGGCTCAGAA 969
Db 1061 GTTCGTGCTATTGAGAAAACCTAGTAAATAACACAGGTATACACATGGCCATGGCTCAGAA 1120
QY 970 ACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1029
Db 1121 ACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1180
QY 1030 TCGTTTACA 1038
Db 1181 TCGTTTACA 1189

RESULT 10
BT006936
LOCUS 1272 bp mRNA linear PRI 13-MAY-2003
DEFINITION Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA, complete cds.
ACCESSION BT006936
VERSION BT006936.1 GI:30582710
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1272)
AUTHORS Kalkine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDs in BD Creator(TM) System Donor vector.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1272)
AUTHORS Kalkine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.

TITLE	Direct Submission
JOURNAL	Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT	This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion (TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ATG' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: http://bioinfo.clontech.com/orfclones .

```

FEATURES             source
Location/Qualifiers
1. .1272
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="BD Creator (TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/notes="Vector: pDNR-Dual"
1. .1272
/codon_start=1
/product="carboxypeptidase B2 (plasma, carboxypeptidase
U)"
/protein_id="AAP35582.1"
/db_xref="GI:30582711"
/translations="MKLSGLAVLPVILFCEQHVAFQSGOVLAAALPTRSQVQVLQI
LTTTIVILVQPTADLIYKKQVHFVNASVDNVKAHLNVSIGPSVLADVEDLI
QQQISNDTIVSPRASYSQYHSLEINYSWIEFTTIRHPDMLTKIHGSSFYKPLYI
LVYSGEQAAKNAIDICGTHAREWISPAFLWFGITFYGIIQGYTNLLRLVDF
VMPVSDVGDYSWKNNRWRKNSRVANNHCITGLDNRNFASKHWCEGASCSGSE
TCGLYPESEPEVKAVASFLRRNIQIKAVISMHSYSOHIVFPEYSYTRSKSKDHELS
LVASAVRALEKTSKNTKTRTHGSGSTLYLAPCGGDDWIYDLGIKYSFTIELRDTGT
GFLVSPRYIKPTCREAPAAVSKIAWIRNV"
373 a 268 c 273 g 358 t
BASE COUNT
ORIGIN

```

Query Match	84.0%	Score	907.4;	DB	9;	Length	1272;	
Best Local Similarity	89.8%	Pred.	No. 5.3e-246;					
Matches 1032; Conservative	0;	Mismatches	6;	Indels	111;	Gaps	1;	
QY	1	ATGAAGCTTTG	CAGCCTTG	CAGTCTT	GTACCAATG	TCTCTCT	TCTGTGAGCAGCATGTC	60
DB	1	ATGAAGCTTTG	CAGCCTTG	CAGTCTT	GTACCAATG	TCTCTCT	TCTGTGAGCAGCATGTC	60
QY	61	TTCCGGTTCC	AGAGTGGCC	AAAGTTT	CTAGCTGCT	TTCCTT	CCTAGAACCTCTAGGCAAGTTCAA	120
DB	61	TTCCGGTTTC	CAGAGTGGCC	AAAGTTT	CTAGCTGCT	TTCCTT	CCTAGAACCTCTAGGCAAGTTCAA	120
QY	121	GTTCTACAGA	AATCTT	ACTACAA	CATATGAGAT	TGTTCT	CTGCGCAGCCGGTAAACAGCTGAC	180
DB	121	GTTCTACAGAN	ATCTT	ACTACAA	CATATGAGAT	TGTTCT	CTGCGCAGCCGGTAAACAGCTGAC	180
QY	181	CTTTATTGTG	AGAAAAA	AAACAAAGT	CCATTTTT	TTTTTGT	TAAATGCATCTGATGCGCAATGTG	240
DB	181	CTTTATTGTG	AGAAAAA	AAACAAAGT	CCATTTTT	TTTTTGT	TAAATGCATCTGATGCGCAATGTG	240
QY	241	AAAGCCCATTT	AAATGTG	AGCGGAAT	TCCATGCAGT	GTCTTGT	CTGCGCAGACGTGGGAAGAT	300
DB	241	AAAGCCCATTT	AAATGTG	AGCGGAAT	TCCATGCAGT	GTCTTGT	CTGCGCAGACGTGGGAAGAT	300
QY	301	CTTTATTCAAC	AGCAGATT	TTC	CAACGACAC	ACAGT	CAGCCCCCGAGCCTCGTACTAT	360
DB	301	CTTTATTCAAC	AGCAGATT	TTC	CAACGACAC	ACAGT	CAGCCCCCGAGCCTCGTACTAT	360
QY	361	GRACAGTAC	CAC	TCAATGAAAT	CTATTCT	TGGATAGA	AAATTTAACTGAGAGGCAT	420
DB	361	GAACAGTAT	CAC	TCAATGAAAT	CTATTCT	TGGATAGA	AAATTTAACTGAGAGGCAT	420
QY	421	CTTGATATG	CTTACAAAAA	TCCACAT	TGGATCCT	CAATTTG	AGAAAGTACCCACTCTATGTT	480

[illegible]

Phelan, M. and Farmer, A.
Direct Submission
Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: <http://bioinfo.clontech.com/orfclones>.
FEATURES
source
1. .1272
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00156L1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/notes="Vector: pDNR-Dual"
1. .>1272
/note="Mutations: 1271:Stop->Leu"
/codon_start=1
/transl_table=1
/product="Homo sapiens carboxypeptidase B2 (plasma,
carboxypeptidase U)"
/protein_id="AAP36662.1"
/db_xref="GI:30584819"
/translation="MKLCSLAVLPIVLFCQHVFAFQSGVLAALPRTSRQVQLQN
LTTTEILVQPVADLVKKQHFVNASVDNKAHLNVSIGPCSVLLADVEDLI
QQQISNDTVSPRASYYEQYSLNIYISWIEFIERHPDMLTKIHGSSFEKYPYLV
LVKVGKEQAKNAIWDGTHAREWISPAFCWFIQHTFYGIIGIYTNLLLVDFY
VMPVNVGDYDYSKNNRMRKRSFVANNHCIGTDLNRNFASKWCEGASSSSCE
TCGLYPESEPEKAVASFLRNINOIKAVISMHSYSHLVFPYSYTRSKSDHELS
LVASEAVRALEKTSKTRTHGHSETLVLPQGGDDWIYDLGKISFTIELRDTGT
GFLPERYIKPTCREAFVAAVSKIAHWIRNVL"
BASE COUNT 372 a 268 c 273 g 359 t
ORIGIN
Query Match 84.0%; Score 907.4; DB 12; Length 1272;
Best Local Similarity 89.8%; Pred. No. 5.3e-246;
Matches 1032; Conservative 0; Mismatches 6; Indels 111; Gaps 1;
QY 1 ATGAAGCTTTGCAGCTTGCAGCTCTTGATCCCATTTCTCTCTGTCGACGACATGTC 60
DB 1 ATGAAGCTTTGCAGCTTGCAGCTCTTGATCCCATTTCTCTCTGTCGACGACATGTC 60
QY 61 TTCGCGTTCAGAGTGCCCAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
DB 61 TTCGCGTTCAGAGTGCCCAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
QY 121 GTTCTACAGATCTTACTACAAATATGATGTTCTCTGCGAGCGGTAAGAGCTGAC 180
DB 121 GTTCTACAGATCTTACTACAAATATGATGTTCTCTGCGAGCGGTAAGAGCTGAC 180
QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAAATGTG 240
DB 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAAATGTG 240
QY 241 AAAGCCCATTTAATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTGGAAGAT 300
DB 241 AAAGCCCATTTAATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTGGAAGAT 300
QY 301 CTTATTCAACAGCAGATTTCCACGACACAGTACGCCCGGCGCTCCGATCGTACTAT 360
DB 301 CTTATTCAACAGCAGATTTCCACGACACAGTACGCCCGGCGCTCCGATCGTACTAT 360
QY 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTTATACTGAGAGCAT 420
|||||

Db 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTTATACTGAGAGCAT 420
QY 421 CCTGATATGCTTTACAAAATCCACATGATGATCTCTCATTTTGGAGAGTACCCACTCTATGTT 480
DB 421 CCTGATATGCTTTACAAAATCCACATGATGATCTCTCATTTTGGAGAGTACCCACTCTATGTT 480
QY 481 TTAAGAGTTTCTGGAAGAAACAAACAGACCCAAAATGCCATATGGATTGACTGTGGAATC 540
DB 481 TTAAGAGTTTCTGGAAGAAACAAACAGACCCAAAATGCCATATGGATTGACTGTGGAATC 540
QY 541 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTGTATAGGCCAT----- 591
DB 541 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTGTATAGGCCATATAAATCAA 600
QY 592 ----- 591
DB 601 TTCTATGGGATATAGGGCAATATATCAATCTCTGAGGCTTGTGGAATTTCTATGTTATG 660
QY 592 -----AATCGAATGTGGAGAAAG 609
DB 661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAGAAATCGAATGTGGAGAAAG 720
QY 610 AACCGTTTCTTCTATCGAACATATTCATTCATCGAACAGACCTGATAGCAACTTTGTC 669
DB 721 AACCGTTTCTTCTATCGAACATATTCATTCATCGAACAGACCTGATAGCAACTTTGTC 780
QY 670 TCCAAACACTGTTGTCGAGGAAGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 729
DB 781 TCCAAACACTGTTGTCGAGGAAGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 840
QY 730 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGAGTGGCTAGTTCTTGTGAGAAATATC 789
DB 841 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGAGTGGCTAGTTCTTGTGAGAAATATC 900
QY 790 AACCGATTAAAGCATACATCAGCATGCTATCTATCCAGCATATAGTGTTCATAT 849
DB 901 AACCGATTAAAGCATACATCAGCATGCTATCTATCCAGCATATAGTGTTCATAT 960
QY 850 TCCATACAGGAATTAAGCAAGACCATGAGGAAGTCTCTCTAGTACGAGTGAAGCA 909
DB 961 TCCATACAGGAATTAAGCAAGACCATGAGGAAGTCTCTCTAGTACGAGTGAAGCA 1020
QY 910 GTTCGTGCTATTGACAAAATAGTAAAAATACCAAGTATACATGCTGCTGCTCAGAA 969
DB 1021 GTTCGTGCTATTGAGAAAATAGTAAAAATACCAAGTATACATGCTGCTCAGAA 1080
QY 970 ACCTTATACCTAGCTCCTGAGGTGGGACGATTTGATGATTTGGGCAATATAT 1029
DB 1081 ACCTTATACCTAGCTCCTGAGGTGGGACGATTTGATGATTTGGGCAATATAT 1140
QY 1030 TCGTTTACA 1038
DB 1141 TCGTTTACA 1149
RESULT 12
BC007057 1715 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
DEFINITION IMAGE:3934520, mRNA, complete cds.
ACCESSION BC007057 GI:13937896
VERSION BC007057
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1715)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK
COMMENT

USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-ehgc.stanford.edu>
Contact: (Dickson, Mark) mdcpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 16 Row: f Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4503004.
Location/Qualifiers
1. 1715
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:1361"
/db_xref="taxon:9606"
/clone="MGC:12495 IMAGE:3934520"
/tissue_type="Skeletal Muscle"
/clone_lib="NIH MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
25. 1296
/codon_start=1
/product="carboxypeptidase B2 (plasma)"
/protein_id="AAH07057.1"
/db_xref="GI:13937897"
/translation="MKLSLAVLPVILFCEQHVFAFQSGVLAALPRTSRQVQLN
LITTYEIVQPVTDLI VKKQVHFVNASVDVNVKAHLNVSGIPCSVLLADVDLI
QQISNDTSVRASASYBOYSLNIYEWIEFIERHPDMLTKIHGSSFEKPYLY
LKVSGKEAKNAIWDICGIHAREWISPAFCWFIQITQFYIGIQYTNLLRLVDFY
VMEVNVQDYDSWKKNRWNRFSYANNHCITGDLNRFNPKSWCEBAGSSCSSE
TYGLYPESEPEVKAVSLRNINOIKAVISMHSYSOHIIVPYSVTRSKSDHRELS
LVASEAVRALEKTSKNTVTHGSETLVLPQGGDDMIYDLGIKISFIELRDTGY
GFLPRLRYIKPTCREAPAAVSKIAWIRNV"

BASE COUNT 518 a 348 c 341 g 508 t

ORIGIN

Query Match 84.0%; Score 907.4; DB 9; Length 1715;
Best Local Similarity 89.8%; Pred. No. 5.4e-246;
Matches 1032; Conservative 0; Mismatches 6; Indels 111; Gaps 1;

QY 1 ATGAAGCTTTGCAGCCTTGAGTCCCTGTGACCAATGTTCTCTCTGTGAGCAGCATGTC 60
DB 25 ATGAAGCTTTGCAGCCTTGAGTCCCTGTGACCAATGTTCTCTCTGTGAGCAGCATGTC 84
QY 61 TTGCGGTTTCAGAGTGCCGAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 120
DB 85 TTGCGGTTTCAGAGTGCCGAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 144
QY 121 GTTCTCAGAAATCTTACTCAACATAGATTGTTCTCTGCGCAGCGGTAAACAGCTGAC 180
DB 145 GTTCTCAGAAATCTTACTCAACATAGATTGTTCTCTGCGCAGCGGTAAACAGCTGAC 204
QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAATGTG 240
DB 205 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAATGTG 264
QY 241 AAGGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGCAGCGTGAAGAT 300
DB 265 AAGGCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGCGCAGCGTGAAGAT 324
QY 301 CTTATTTCACAGCAGATTTTCACAGCAGCAGTCAAGCCCCGAGCCCTCCGATCGTACTAT 360
DB 325 CTTATTTCACAGCAGATTTTCACAGCAGCAGTCAAGCCCCGAGCCCTCCGATCGTACTAT 384

FEATURES
source

CDS

QY 361 GAACAGTATCAGTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 420
DB 385 GAACAGTATCAGTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 444
QY 421 CTTGATATCTTACAAAATCCACATTTGATCTCTCAATTTGAGAGTACCCACTCTATGTT 480
DB 445 CTTGATATCTTACAAAATCCACATTTGATCTCTCAATTTGAGAGTACCCACTCTATGTT 504
QY 481 TTAAGGTTTCTGGAAGAAGAAACAAACAGACCAAAAATGCCATATGGAATTTGACTGTGAATC 540
DB 505 TTAAGGTTTCTGGAAGAAGAAACAAACAGACCAAAAATGCCATATGGAATTTGACTGTGAATC 564
QY 541 CATGCCAGAGATGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 591
DB 565 CATGCCAGAGATGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 624
QY 592 ----- 591
DB 625 TTCTATGGGATATATAGGGCAATATACCAATCTCTCTGAGGCTTGTGGATTTCTATGTTATG 684
QY 592 ----- 591
DB 685 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAGAAAGATCGAATGTGGAGAAAG 744
QY 610 AACCGTTCTTTCTATGCGAACCAATCATTCGATCGGAACAGACCTGAATAGCAACTTTGTC 669
DB 745 AACCGTTCTTTCTATGCGAACCAATCATTCGATCGGAACAGACCTGAATAGCAACTTTGTC 804
QY 670 TCCAAACACTGGTGTGAGGAAGTGATCCAGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 729
DB 805 TCCAAACACTGGTGTGAGGAAGTGATCCAGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 864
QY 730 CTTTATCTCTGAGTCAGAACACAGAAAGTGAAGGAGTGGCTAGTTCTCTTGAAGAAATATC 789
DB 865 CTTTATCTCTGAGTCAGAACACAGAAAGTGAAGGAGTGGCTAGTTCTCTTGAAGAAATATC 924
QY 790 AACAGGATTTAAAGCATACATCATGATGATTCATCTCTCCAGCATATAGTGTGTTTCCATAT 849
DB 925 AACAGGATTTAAAGCATACATCATGATGATTCATCTCTCCAGCATATAGTGTGTTTCCATAT 984
QY 850 TCCATATACAGGAAGTAAAGCAACCAATGAGGAATCTGTCTAGTAGCCAGTGAAGCA 909
DB 985 TCCATATACAGGAAGTAAAGCAACCAATGAGGAATCTGTCTCTAGTAGCCAGTGAAGCA 1044
QY 910 GTTCGTGCTATTGACAAAATAGTAAATAATACAGGTATACATACATGCGCATGCTCAGAA 969
DB 1045 GTTCGTGCTATTGAGAAAATAGTAAATAATACAGGTATACATGCGCATGCTCAGAA 1104
QY 970 ACCTTATACCTAGCTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCATCAATAT 1029
DB 1105 ACCTTATACCTAGCTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCATCAATAT 1164
QY 1030 TCGTTTACA 1038
DB 1165 TCGTTTACA 1173

RESULT 13
AB021968
LOCUS
DEFINITION
AB021968
ACCESSION
VERSION
AB021968.1 GI:9558447
KEYWORDS
SOURCE
MUS MUSCULUS (HOUSE MOUSE)
ORGANISM
MUS MUSCULUS
REFERENCE
1 (sites)
AUTHORS
Sato, T., Miwa, T., Akatsu, H., Matsukawa, N., Obata, K., Okada, N.,
Campbell, W. and Okada, H.
TITLE
Pro-carboxypeptidase R is an acute phase protein in the mouse.

AB021968
MUS MUSCULUS mRNA for carboxypeptidase R, complete cds.
1421 bp mRNA linear ROD 27-JUL-2000

whereas carboxypeptidase N is not
J. Immunol. 165 (2), 1053-1058 (2000)
20341711
PUBMED 10878383
REFERENCE 2 (bases 1 to 1421)
AUTHORS Sato, T.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1998) Tomoo Sato, Fukushima Hospital, Chofu
Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi
441-8124, Japan (E-mail: tomoo@tcp-ip.or.jp, Tel: 81-532-46-7511,
Fax: 81-532-46-4899)
FEATURES
Location/Qualifiers
1. 1421
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
17. 1285
/codon_start=1
/product="carboxypeptidase R"
/protein_id="BA03402.1"
/db_xref="GI:9558448"
/translation="MKLHGLILVAILVEHQGFQSGQVLSPALPRTSRQVOLLQNL
TTTTYVLPQVTAETIEKKKEVFFVNASDVDSVKAHLNVSRIPIFNLMNVNEDIE
QOTFNEDPSRASAYEYQHSLEIYSWIEVTEHQPMLOKLYLSSPEKYPFLVFL
KVGSKQRIKNAIWDGHIHAREWISPAFLWIFGVYVTOFHGENLYTLRLHVDYFI
MPVNVVDGYDYWKKNMWEKNSAHKNRCVGTDLNRNPFASKHWCSEKSSCSET
YCGLPESBEVKAADVFLNRIDHIAKAYISMHSYSQOILPPYSYNRSKSDHELSL
VASEAVRALESINKNTRYTHSGSLESYLAPGSSDDWIYDLGIKYSFTIELRDTGRYG
FLPERYIKPTCEALAAASKIVMHIRNT"
1421
polyA_site
BASE COUNT 418 a 320 c 305 g 378 t
ORIGIN
Query Match 59.6%; Score 643.4; DB 10; Length 1421;
Best Local Similarity 76.1%; Pred. No. 4e-171;
Matches 874; Conservative 0; Mismatches 161; Indels 114; Gaps 2;
QY 1 ATGAGCTTTGCAGCTTGCAGCTTGTACCTGTTGACCATGTTCTCTCTGTGAGCAGCATGTC 60
DB 17 ATGAGCTTTCATGCGCTTGAATCCTGTTGAGCCA---TCATCTCTATGAGCAGCATGGC 73
QY 61 TTGCGCTTCCAGAGTGCCCAAGTTCTAGCTGCTCTCTTAGAACCTCTAGGCAAGTTCAA 120
DB 74 TTGCGCTTCCAGAGTGCCCAAGTTTATCTGCTCTTCCAGAACCTCTCAGGCAAGTTCAA 133
QY 121 GTTCTACAGAACTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 180
DB 134 CTACTTCAGAACTTACTACACATGATGAGTCTGTTCTCTGGCAGCGGTAACAGCTGAA 193
QY 181 CTTATTGTGAAGAAAACCAAGTCCATTTTTTTGTAATGCATCTGATGTCGAAATGTG 240
DB 194 TTCATCAGAGAAGAAAAGGAAGTCCACTTTTTTTGTAATGCATCTGATGTCGAGTGTG 253
QY 241 AAAGCCATTTAAATGTGAGCGGAATTCATGATGCTGTTCTGTCGAGCGGTAAGAT 300
DB 254 AAAGCGATTTAAATGTGAGCAGAAATTCATTTAACTGATGAACAGTGGAGGAC 313
QY 301 CTTATTCAACAGCAGATTTTCCAGCAGACAGTACAGCCCCGAGCCTCCGATCGTACTAT 360
DB 314 CTAATTGAACAGCAGACTTTTCAATGACACAGGTGAGCCCCGCGCTCCGCTTCACTAT 373
QY 361 GAACAGTATCACTCAATAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 420
DB 374 GAGCAGTATCACTCGCTTAAATGAAATCTATTCTTGGATAGAAATGATCACTGAACGAT 433
QY 421 CTTGATATGCTTCAAAAAATCCCATTTGATGATCTCTTTCAGAAAGTACCCACTCTATGT 480
DB 434 CTTGATATGCTTCGAAAAATCTCATCGGATCATCTCAGAAAGTACCCACTTTATGT 493
QY 481 TTAAGGTTTTCTGAAAAGAACCAACAGCCAAAAATGCCATATGATGATGCTGTGGAATC 540
DB 494 TTAAGGTTCTCAGGAAGGAACCAAGAAATCAAAAAATGCCATCTGATCGACTGTGGAATC 553

QY 541 CATGCCAGAGAATGGATCTCTCTGCTTTCTGCTTTGTTGTTTCATAGGC----- 588
DB 554 CATGCCAGAGAATGGATTTTCACTGCTTTTCTGTTTGGTTTCATAGGTACGTGACAA 613
QY 589 ----- 588
DB 614 TTCCATGGGAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTCTTACATCATG 673
QY 589 -----CATATCGAATGTGGAGAAAG 609
DB 674 CCGGTGATGAACGTGGATGGCTATGACTACACGTGGAAGAAATCGAATGTGGAGGAG 733
QY 610 AACCGTTCTTTCTATCGGACAAATCATTCATCGGACAGACCTGATAGTACCTTTTGTGTC 669
DB 734 AACCGTCTGCTCACAAGAACCAACCGCTGGTGGGACAGACCTGAAACAGGAACTTCGCT 793
QY 670 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATGCTCGGAAACCTTACTGTGGA 729
DB 794 TCCAAACACTGGTGTGAGGAAGTGCATCAAGTTCTCTCTGCTCTGAACCTTACTGTGGA 853
QY 730 CTTTATCTCGAGTCAGAACACAGAAAGTGAAGGCGAGTGGTAGTTTCTTGTGAGAAATATC 789
DB 854 CTTTATCTCGAGTCTGAGCCAGAGGTGAAGGCGAGTGGTAGTTTCTTGTGAGAAATATC 913
QY 790 AACCAATTAAGCAGATCATGACATGCTTATCTATCTCCAGCATATAGTGTGTTTCCATAT 849
DB 914 GACCAATTAAGCTTACATGACTGCTATCTATCTCCCAACAAATATCTGTTTCCCTAT 973
QY 850 TCCATATACAGCAAGTAAAGCAAGACCATGAGGAACCTGTCTAGTAGCCAGTGAAGCA 909
DB 974 TCCATATACAGCAAGCAAGCAAGGACACAGGAACCTGTCTAGTGGCCAGCAAGCA 1033
QY 910 GTTCTGCTATTGACAAAACCTAGTAAAAATACAGGTATATACATGCGCATGCTCAGAA 969
DB 1034 GTTCTGCTAAATGAAAGTATTAATAAAAAACACCCAGGTATACACACGCGAGTGGCTCAGAA 1093
QY 970 ACCTTATATCTAGCTCTCGAGTGGGAGGATTTGGATCTATGATTTGGGCATCAATAT 1029
DB 1094 AGTTTATATCTAGCTCTCGAGGTTCTGAGGATTTGATGATTTGGGCATCAATAT 1153
QY 1030 TCGTTTACA 1038
DB 1154 TCGTTTACA 1162
RESULT 14
AF164524
LOCUS
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)
ACCESSION AF164524
VERSION AF164524.1 GI:7416966
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1490)
AUTHORS Marx, P.F., Wagenaar, G.T., Reijerkerk, A., Tiekstra, M.J., van Rossum, A.G., Gebbink, M.F. and Meijers, J.C.
TITLE Characterization of mouse thrombin-activatable fibrinolysis inhibitor
JOURNAL Thromb. Haemost. 83 (2), 297-303 (2000)
MEDLINE 20201996
PUBMED 10739389
REFERENCE 2 (bases 1 to 1490)
AUTHORS Marx, P.F., Wagenaar, G.T.M., van Rossum, A.G.S.H. and Meijers, J.C.M.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100,
Utrecht 3584 CX, The Netherlands
FEATURES
Location/Qualifiers
1. 1490
source

```
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:477035"
1..1490
/gene="Tafi"
84..1352
CDS
/gene="Tafi"
/function="carboxypeptidase"
/codon_start=1
/product="thrombin-activatable fibrinolysis inhibitor"
/product_id="AA62385.1"
/db_xref="GI:7416967"
/translations="MKHLGLILVAILLEQHFQFQSGQVLSPALPRTSRQVOLLQNL
TTTVEVLQPVFAETIEKKKVEHFFVNASDSDVSQAHNLNVSRIPNFLMNVNVEDLIE
QOTFNDFPRASAYEQHSLNEIYSEIETEQHDMQIQYIGSSFEKPYPLVL
KVSQKQRIKNAIWDGCIHAREWIFPAFLWIFGVTFQHGKNIYTLRLHVDYFI
MPVMVDGYDYTWKRNMRKNSAHKNNRCVGTDLNRFASFHKWCEKASSSCSET
YCLYPESEPEVKAVADFLRNIDHIIKAYISMHSYSQQILPFPYSYNRSKDHEELSL
VASEARPEIESINKNRYTHSGSESILAPGSSDDWIYDLGIKYSFTIELRDTGRYG
FLPPIRYIKPTCAEALAAISKIVMVRNT"
BASE COUNT 438 a 337 c 322 g 393 t
ORIGIN

Query Match 59.6%; Score 643.4; DB 10; Length 1490;
Best Local Similarity 76.1%; Pred. No. 4e-171;
Matches 874; Conservative 0; Mismatches 161; Indels 114; Gaps 2;

QY 1 ATGAAGCTTTGCAGCTTGCAGTCTTGATCCCATTTGCTCTCTCTGTGAGCAGCATGTC 60
DB 84 ATGAAGCTTTCATGCGCTTGGAATCTCTGTAGCCA---TCATCTCTATGAGCAGCATGGC 140
QY 61 TTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
DB 141 TTCGCGTTTCAGAGTGCCCAAGTTTATCTGCTCTTCCAGAACCTCCAGGCAAGTTCAA 200
QY 121 GTTCTACAGAACTTACTACACATATGAGATTGTTCTCTGCGCAGCGGTTAAACAGCTGAC 180
DB 201 CTACTTTCAAGATCTTACTACAGATGATGAGTGCTGTTCTCTGCGCAGCGGTTAAACAGCTGAA 260
QY 181 CTTATTGTGAAGAAAACCAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 240
DB 261 TTCTATGAGAAAGAAAGGAAGTCCACTTTTGTGATCGCTCTGATGTCGACAGTGC 320
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTTGTCTGCGCAGCAGTGGGAAGAT 300
DB 321 AAAGCCCATTTAAATGTGAGCAGAAATTCATTTAAAGTCTGATGAACACAGTGGAGGAC 360
QY 301 CTTATTCAACAGCAGATTTCCAGACACAGTCAAGCCCCCGAGCGCTCCCGCATCGTACTAT 360
DB 381 CTAATTGAACAGCAGACTTTCAATGACACGCTCAGCCCCCGCGCTCCCGCTTCATACTAT 440
QY 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATTAACGTGAGAGCAT 420
DB 441 GAGCAGTATCACTCGCTTAAATGAATCTATTCTCTGGATAGAAATTTATTAACGTGAGAGCAT 500
QY 421 CTTGATATGCTTCAAAAATCCCAATGGATCTCTATTTCAGAAAGTACCCACTCTATGTT 480
DB 501 CTTGATATGCTTCCAGAAATCTACATCGGATCATCTTCAGAAAGTACCCACTTATGTT 560
QY 481 TTAAGGTTTCTGGAAGAAACAAACAGCAGCAAAATGCCATATGATTTGATCTGGAATC 540
DB 561 TTAAGGTTCTCAGAAAGAAACAAAGAAATCAAAAATGCCATCTGGATCGACTGTGGAATC 620
QY 541 CATGCCAGAAATGGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 588
DB 621 CATGCCAGAAATGGATTTTCACTGCTTCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 680
QY 589 ----- 588
DB 681 TTCCATGGGAAAGAAATCTGTATACAGACTTCTGAGGACAGTGGATTTCTATCATG 740
QY 589 -----CATATCGAATGTGGAAAG 609
```

/translation= "MKLYGLGLVLAIIIEYKHLGAFQSHVLSALPRTSRQVOLLQWL
TTTVEVLPWPVTEAFTEKKKEHFFVNADSVNSVXAYLNASRI PFNLNMNVEDLIQ
QOVSNDTVSPASSYVEOYHSNEIYVSWIEVTEQHPDMLKLIYGSSEYKXPLVYL
KVSQKEHRVKNALWIDCGIHAREWISPAFLWIFVYQPHKENIYIRLLRHVDVFI
MPWNVDGYDYTWKKNRMWRKNSVHNNRCVGTDLNRFNFAKHWCEKGAASSPSCSE
YCGLYPESEPEKAVADFLRNINHIKAYISMSHSYQQILFPYSYNSRKSKEEELSL
VASEAVRAIESINKNTRYTHSGSESLYLAPEGSDDWIYDLGIKYSFTIELRDTGRYG
FLLPERFIKPTCEALAAVSKIAHWIRNS"
BASE COUNT 417 a 334 c 309 g 365 t
ORIGIN

Query Match 59.5%; Score 642.4; DB 10; Length 1425;
Best Local Similarity 76.0%; Pred. No. 7.6e-171;
Matches 873; Conservative 0; Mismatches 161; Indels 114; Gaps 2;

Qy	1	ATGAAGCTTTGCAGCCTTGAGTCTTGATACCAATGTTCTCTCTGTGTGAGCAGCATGTC	60
Db	100	ATGAAGCTTTTATGGCCTTGGAGTCTCTGTGAGCA---TCATCTCTATGAGAAGCATGCG	156
Qy	61	TTGCGGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA	120
Db	157	CTTGCCCTTCAGAGTGCCCATGTTCTATCTGCTCTCTGACCTCCAGGCAAGTTCAA	216
Qy	121	GTTCTACAGAACTTACTCAACAATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC	180
Db	217	CTTCTTCAGAACTCTCACTCAAACTTACGAGTTGTTCTCTGGCAGCCAGTGACAGCTGAA	276
Qy	181	CTTATTGTGAAGAAAAAACAAGTCCATTTTTTGTAAATGATCTGATCGACAATGTG	240
Db	277	TTCAATTGAGAAGAAAAAGGAAGTCCACTTCTTTGTGAATGCTGCTGATGTCAACAGTGC	336
Qy	241	AAAGCCCATTTAAATGTAGCGGAATTCATGACAGTCTTGTGGCAGACGTTGGAGAT	300
Db	337	AAAGCCCATTTAAATGTAGCGAGAAATTCATTTACGCTCTGTGATGAACAACGTTGGAGAT	396
Qy	301	CTTATTCAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAGCCTCCGCATCGTACTAT	360
Db	397	CTAATTCAACAGCAGACGTCCTCAATGACACTGTTAGCCCGGAGCCTCTCTCTCATACTAT	456
Qy	361	GACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATACTGAGAGCAT	420
Db	457	GACAGTATCACTCGTTTAAATGAAATCTATTCTCTGGATGAAGTTATACTGAACAGCAC	516
Qy	421	CCTGATATGCTTACAAAATCCCATTTGGATGCTCTATTTCAGAGTACCCACTCTATGTT	480
Db	517	CCTGACATGCTCCAGAAAATCTCATTTGGATGCTCTATGAAAGTACCACCTTTATGTG	576
Qy	481	TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAATGCCATATGGATTGACTGTGGAATC	540
Db	577	TTAAAGGTTCTCAGAAAAGGAACACAGAGTCAAAAATGCCATATGGATCGACTGTGGAATC	636
Qy	541	CATGCCAGAGAAATGATCTCTCTGCTTCTCTGCTTGTGTTGTTATAGGCCAT-----	591
Db	637	CATGCCAGAGATGGAATTTACACAGCTTTCTGCTTGTGTTTCTATAGGCTATGTAACGCAA	696
Qy	592	-----	591
Db	697	TTCCATGGGAAGAAAATACATACACAGACTTCTGAGGCACGTTGGATTCTCATTTATG	756
Qy	592	-----AATCGAATGTGGAGAAAG	609
Db	757	CCAGTGATGAATGTGACGGCTACGACTACACGTTGGAAAAGAAATCGAATGTGGAGAAAG	816
Qy	610	AACGGTTCTTCTATGGAACAATCATTTGCATCGGAACAGACCTGGAATAGCACTTTGTC	669
Db	817	AACCGCTCTGTCCACATGAACAACCCGCTGGTGGGCAACAGACCTGGAACAGGAACTTCGCT	876
Qy	670	TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA	729
Db	877	TCCAAACACTGGTGTGAGGAGGGGCATCAAGTTTCTCTGCTCTGAGACCTACTGTGGA	936
Qy	730	CTTTATCTCTGAGTCAGAACACAGAGTGAAGGCACTAGTTTCTTTGAGAAGAAATATC	789

Search completed: January 1, 2004, 13:55:55
Job time : 4083.16 secs

Db	937	CTTTACCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTCTGAGGAGAAATATC	996
Qy	790	AACCAATTAAAGCATACATCAGCATGCCATTCTACTATCCAGCATATAGTTTCCATAT	849
Db	997	AACCAATTAAAGCTTACATCAGTATGCACTCATCTCCAGCAAAATACTGTTTCCCTAT	1056
Qy	850	TCCTATACACGAAGTAAAGCAAGAACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCA	909
Db	1057	TCCTACACAGAAGCAAGCAAGGACCAAGGAACTGTCTCTAGTGGCCAGGGAAGCA	1116
Qy	910	GTTTCGTGCTATTTCACAAAACCTAGTAAAAATACCAAGGTATACATATGCTCCATGCTCAGAA	969
Db	1117	GTTTCGTGCCATTGAAAAGTATTAATAAAAAACACCAAGGTACACATGCGCAGTGGCTCAGAA	1176
Qy	970	ACCTTATACCTAGCTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT	1029
Db	1177	AGTTTATATCTAGCTCTCTGAGGTTCGTATGATTTGGATCTATGATTTGGGCATCAAAATAT	1236
Qy	1030	TCGTTTAC 1037	
Db	1237	TCGTTTAC 1244	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:40:19 ; Search time 317.12 Seconds
(without alignments)
9193.336 Million cell updates/sec

Title: US-09-980-881a-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagcttgacgcttgc.....tttgcgctgtctctaaaa 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1080	100.0	1573	21	AAC81962 Human brain carbox
2	1025.2	94.9	1625	25	ABX13670 Human protease cDN
3	909	84.2	1272	20	AAV74302 Human plasma carbo
4	909	84.2	1728	24	ABN95899 Gene #2397 used to
5	909	84.2	1749	17	AAV11671 Human plasma carbo
6	909	84.2	1749	18	AAV62846 Human plasma carbo
7	907.4	84.0	1749	14	AAQ41001 Human plasma carbo
8	441	40.8	1400	21	AAV18005 Lung cancer associ

9	331.2	30.7	416	25	ABX49430	Bovine EST associa
10	282.4	26.1	397	25	ABX43260	Bovine EST associa
11	204	18.9	55827	25	ABX13671	Human protease gen
12	167	15.5	231	24	ABN19133	Human ORFX polynuc
13	129.6	12.0	431	24	ABK62561	Rat sequence diffe
14	124	11.5	927	17	AAV35760	Rat mature carboxy
15	123.4	11.4	921	16	AAQ90601	Porcine carboxypep
16	123.4	11.4	1215	16	AAQ90600	Porcine Tyr-His-Me
17	109	10.1	1263	19	AAV41795	Human pancreatic c
18	107.6	10.0	2023	25	ABZ75116	Anti-human seminal
19	107.4	9.9	999	17	AAV42500	Mature HCPB coding
20	107.4	9.9	1053	17	AAV42497	mature HCPB- (His)6
21	107.4	9.9	1263	17	AAV42494	Human pancreatic c
22	107.4	9.9	1284	17	AAV42506	ProHCPB gene with
23	104.2	9.6	1053	20	AAV424804	Human carboxypepta
24	104.2	9.6	1059	17	AAV42511	Modified HCPB (D25
25	104.2	9.6	1059	17	AAV42512	Modified HCPB (D25
26	104.2	9.6	1059	18	AAV62787	Carboxypeptidase B
27	104.2	9.6	1059	18	AAV62788	Carboxypeptidase B
28	102.6	9.5	1059	18	AAV62805	Carboxypeptidase B
29	102.6	9.5	1059	18	AAV62806	Carboxypeptidase B
30	102.6	9.5	1059	18	AAV62791	Carboxypeptidase B
31	102.6	9.5	1059	18	AAV62801	Carboxypeptidase B
32	101	9.4	1059	18	AAV62789	Carboxypeptidase B
33	101	9.4	1059	18	AAV62790	Carboxypeptidase B
34	101	9.4	1870	18	AAV17322	DNA encoding human
35	101	9.4	1870	19	AAV41819	Human carboxypepti
36	101	9.4	2154	18	AAV17331	PreproHCPB-linker-
37	99.4	9.2	1059	18	AAV62808	Carboxypeptidase B
38	99.4	9.2	1059	18	AAV62807	Carboxypeptidase B
39	99.2	9.2	1059	18	AAV62800	Carboxypeptidase B
40	97.8	9.1	1059	18	AAV62807	Carboxypeptidase B
41	97.8	9.1	1059	18	AAV62803	Carboxypeptidase B
42	96.2	8.9	1059	18	AAV62802	Carboxypeptidase B
43	94.6	8.8	1332	24	ABL65831	Lung cancer relate
44	92.8	8.6	1302	25	AAV77254	Nucleotide sequenc
45	92.8	8.6	1311	22	AAV03837	Human carboxypepti

ALIGNMENTS

RESULT 1
AAC81962
ID AAC81962 standard; cDNA; 1573 BP.
XX AAC81962;
AC AAC81962;
XX 01-MAR-2001 (first entry)
DT Human brain carboxypeptidase B cDNA.
DE Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 18..1100
FT /*tag= a
FT /product= "carboxypeptidase B"
XX WO200066717-A1.
PN 09-NOV-2000.
PD 01-MAY-2000; 2000WO-JP02878.
PF 30-APR-1999; 99JP-0125169.
PR XX

XX WPI; 2003-102387/09.
DR P-PSDB; ABG72093, ABG72095.
XX
PT New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
PT
XX
PS Claim 4; Fig 1; 80pp; English.
XX
XX The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for
CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorder. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC cDNA, the gene for which is located on chromosome 13.
XX
SQ Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;

Query Match 94.9%; Score 1025.2; DB 25; Length 1625;
Best Local Similarity 99.2%; Pred. No. 8.5e-311;
Matches 1030; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGAAGCTTTCAGAGTCCAGTCCCTGTGACCCATGTTCTCTCTCTGTGAGCAGCATGTC 60
DB |||||
QY 17 ATGAAGCTTTCAGAGTCCAGTCCCTGTGACCCATGTTCTCTCTCTGTGAGCAGCATGTC 76
DB |||||
QY 61 TTCGCGTTCCAGAGTCCAGTCCCTGTGACCCATGTTCTCTCTCTGTGAGCAGCATGTC 120
DB |||||
QY 77 TTCGCGTTCCAGAGTCCAGTCCCTGTGACCCATGTTCTCTCTCTGTGAGCAGCATGTC 136
DB |||||
QY 121 GTTCTACAGATCTTACTACACATATGAGATGTTCTCTCTGTGAGCAGCATGTC 180
DB |||||
QY 137 GTTCTACAGATCTTACTACACATATGAGATGTTCTCTGTGAGCAGCATGTC 196
DB |||||
QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGAAATGATGATGTCGCAATGTG 240
DB |||||
QY 197 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGAAATGATGATGTCGCAATGTG 256
DB |||||
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGCGAGACGTGGAAGAT 300
DB |||||
QY 257 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGCGAGACGTGGAAGAT 316
DB |||||
QY 301 CTTATTCAACAGCAGATTTTCCACGACACAGTACAGTCCGAGCCCTCCGATCGTACTAT 360
DB |||||
QY 317 CTTATTCAACAGCAGATTTTCCACGACACAGTACAGTCCGAGCCCTCCGATCGTACTAT 376
DB |||||
QY 361 GAACAGTATCACTCACTAAATGAATCTAATCTTGTGATAGAAATTTATACTGAGAGCAT 420
DB |||||
QY 377 GAACAGTATCACTCACTAAATGAATCTAATCTTGTGATAGAAATTTATACTGAGAGCAT 436
DB |||||
QY 421 CCTGATATGCTTACAAAAATCCACATTTGGATCTCTATTTGAGAGTACCCACTTATGTT 480
DB |||||
QY 437 CCTGATATGCTTACAAAAATCCACATTTGGATCTCTATTTGAGAGTACCCACTTATGTT 496
DB |||||

QY 481 TTAAGGTTTCTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 540
DB |||||
QY 497 TTAAGGTTTCTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 556
DB |||||
QY 541 CATGCCAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 600
DB |||||
QY 557 CATGCCAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 616
DB |||||
QY 601 TGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 660
DB |||||
QY 617 TGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 676
DB |||||
QY 661 AACTTTGTCTTCAAAACACTGGTGTGAGGAGGTGATCCAGTTCCTCATCTGCTCGGAAACC 720
DB |||||
QY 677 AACTTTGTCTTCAAAACACTGGTGTGAGGAGGTGATCCAGTTCCTCATCTGCTCGGAAACC 736
DB |||||
QY 721 TACTGTGGACTTTTATCTCTGAGTCAGAACAGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 780
DB |||||
QY 737 TACTGTGGACTTTTATCTCTGAGTCAGAACAGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 796
DB |||||
QY 781 AGAAATATCAACAGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 840
DB |||||
QY 797 AGAAATATCAACAGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAG 856
DB |||||
QY 841 TTTCCATATTTCTTATACAGAAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAG 900
DB |||||
QY 857 TTTCCATATTTCTTATACAGAAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAG 916
DB |||||
QY 901 AGTGAAGCAGTTCGTCTTATGACAAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAG 960
DB |||||
QY 917 AGTGAAGCAGTTCGTCTTATGACAAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAGCAAGTAAAG 976
DB |||||
QY 961 GGCTCAGAAACCTTATACCTAGTCTCTGAGAGTGGGAGCAGTATGGATCTATGATTTGGGC 1020
DB |||||
QY 977 GGCTCAGAAACCTTATACCTAGTCTCTGAGAGTGGGAGCAGTATGGATCTATGATTTGGGC 1036
DB |||||
QY 1021 ATCAAAATATTCGTTTACA 1038
DB |||||
QY 1037 ATCAAAATATTCGTTTACA 1054
DB |||||

RESULT 3
AAV74302
ID AAV74302 standard; cDNA; 1272 BP.
XX AC AAV74302;
XX AC AAV74302;
XX 28-APR-1999 (first entry)
XX Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
DE DE
XX Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW KW polymorphism detection; thrombotic disease; ds.
XX Homo sapiens.
OS OS
XX WO9855645-A1.
XX PN
XX 10-DEC-1998.
XX PD
XX 02-JUN-1998; 98WO-EP03244.
XX PF
XX 03-JUN-1997; 97US-0869057.
XX PR
XX (SCHD) SCHERING AG.
XX PA
XX Morser MJ, Nagashima M;
PI PI
XX WPI; 1999-045800/04.
DR P-PSDB; AAW92270.
XX
XX Detecting new polymorphism of human plasma carboxypeptidase B -

PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Example 1; Page 24; 35pp; English.
XX
CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBhr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphisms of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphisms in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphisms within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
SQ Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

Query Match 84.2%; Score 909; DB 20; Length 1272;
Best Local Similarity 89.9%; Pred. No. 2.2e-274;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;
QY 1 ATGAAGCTTTGACGCTTGGAGTCCCTGAGTCCCTGACCCATTTCTCTCTGAGGAGCATGTC 60
DB 1 ATGAAGCTTTGACGCTTGGAGTCCCTGAGTCCCTGACCCATTTCTCTCTGAGGAGCATGTC 60
QY 61 TTCCGCTTCCAGAGTGGCCAGTGTCTAGTGTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
DB 61 TTCCGCTTCCAGAGTGGCCAGTGTCTAGTGTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
QY 121 GTTCTACAGAACTTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 180
DB 121 GTTCTACAGAACTTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGAC 180
QY 181 CTTATTGTGAGAAAACCAAGTCCATTTTGTAAATGATGATGATGATGATGATGATGATG 240
DB 181 CTTATTGTGAGAAAACCAAGTCCATTTTGTAAATGATGATGATGATGATGATGATGATG 240
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTTGGAAGAT 300
DB 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTTGGAAGAT 300
QY 301 CTTATTCAACAGCAGATTTCAACGACACAGTACGAGCCCGAGCCTCCGATCGTACTAT 360
DB 301 CTTATTCAACAGCAGATTTCAACGACACAGTACGAGCCCGAGCCTCCGATCGTACTAT 360
QY 361 GAACAGTATCACTCACTAATGAAATCTATTCTTGGATAGAAATTTATTAACGTGAGGCGAT 420
DB 361 GAACAGTATCACTCACTAATGAAATCTATTCTTGGATAGAAATTTATTAACGTGAGGCGAT 420
QY 421 CCTGATATGCTTCAAAAATCCCATTTGGATTCCTTCTGAGAGTACCCACTCTATGTT 480
DB 421 CCTGATATGCTTCAAAAATCCCATTTGGATTCCTTCTGAGAGTACCCACTCTATGTT 480
QY 481 TTAAGGTTTCTGAAAAGAACAAACAGCCAAAATGCCATATGATTTGATGATGATGATGATG 540
DB 481 TTAAGGTTTCTGAAAAGAACAAACAGCCAAAATGCCATATGATTTGATGATGATGATGATG 540
QY 541 CATGCCAGAGATGATCTCTCTGCTTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTG 591
DB 541 CATGCCAGAGATGATCTCTCTGCTTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTG 600
QY 592 ----- 591
DB 601 TTCTATGGATAATAGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 660
QY 592 -----AATCAATGTGAGAAAG 609
DB 661 CCGGTGTTAATGTGAGCGGTTATGACTACTCATGAGAAAAGAAATCGAATGTGAGAAAG 720

QY 610 AACCGTTCTTTCTATTCGAAACATCATTCATCGGAACAGACCTGAATAGCAACTTTGTC 669
DB 721 AACCGTTCTTTCTATTCGGAACATCATTCATCGGAACAGACCTGAATAGCAACTTTGTC 780
QY 670 TCCAAACACTGTGTGTGAGGAAGTGCATCCAGTTCCTCATGTCTCGAAACCTTACTGTGGA 729
DB 781 TCCAAACACTGTGTGTGAGGAAGTGCATCCAGTTCCTCATGTCTCGAAACCTTACTGTGGA 840
QY 730 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGAGTGGCTAGTTCTTGTGAGAAATATC 789
DB 841 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGAGTGGCTAGTTCTTGTGAGAAATATC 900
QY 790 AACCAAGTAAAGCATACATCAGCATGCATTCTATCTCCAGCATATAGTGTGTTTCCATAT 849
DB 901 AACCAAGTAAAGCATACATCAGCATGCATTCTATCTCCAGCATATAGTGTGTTTCCATAT 960
QY 850 TCCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGTCTCTAGTACGAGTGAAGCA 909
DB 961 TCCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGTCTCTAGTACGAGTGAAGCA 1020
QY 910 GTTCTGTCTATTGACAAAACCTAGTAAATACAGGTATACATACATGCGCATGCGTCAAG 969
DB 1021 GTTCTGTCTATTGAGAAAACCTAGTAAATACAGGTATACATACATGCGCATGCGTCAAG 1080
QY 970 ACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGGATCTATCATTTTGGCATCAAAATAT 1029
DB 1081 ACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGGATCTATCATTTTGGCATCAAAATAT 1140
QY 1030 TCGTTTACA 1038
DB 1141 TCGTTTACA 1149

RESULT 4

ABN95899
ID ABN95899 standard; DNA; 1728 BP.

XX ABN95899;

DT 13-AUG-2002 (first entry)

XX Gene #2397 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30589.

XX 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample

XX Claim 1; SEQ ID NO 2397; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver

CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatocellular, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;

Query Match 84.2%; Score 909; DB 24; Length 1728;
Best Local Similarity 89.9%; Pred. No. 2.6e-274;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

QY 1 ATGAAGCTTTGACGCTTGAGCTCTTGTTACCAATGTTCTTCTGTGTGAGCAGATGTC 60
DB 20 ATGAAGCTTTGACGCTTGAGCTCTTGTTACCAATGTTCTTCTGTGTGAGCAGATGTC 79
QY 61 TTGCGGTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
DB 80 TTGCGGTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 139
QY 121 GTTCTACAGAACTTACTACAAATAGATGATGTTCTCTGCGAGCGGTAACAGCTGAC 180
DB 140 GTTCTACAGAACTTACTACAAATAGATGATGTTCTCTGCGAGCGGTAACAGCTGAC 199
QY 181 CTATTATGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTCGACAAATGTG 240
DB 200 CTATTATGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTCGACAAATGTG 259
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGCTGGCAGACGTTGGAAGAT 300
DB 260 AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGCTGGCAGACGTTGGAAGAT 319
QY 301 CTATTATCAACAGCAGATTTTCCACGACACAGTCCAGCCCGAGCCCTCCGATCGTACTAT 360
DB 320 CTATTATCAACAGCAGATTTTCCACGACACAGTCCAGCCCGAGCCCTCCGATCGTACTAT 379
QY 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 420
DB 380 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGGCAT 439
QY 421 CTTGATATGCTTACAAAATCCCATTTGGATTCCTCATTTGAGAAAGTACCCACTATGTT 480
DB 440 CTTGATATGCTTACAAAATCCCATTTGGATTCCTCATTTGAGAAAGTACCCACTATGTT 499
QY 481 TTAAGGTTTCTGGAAGAACAAACAGCCAAATGCCATATGATGATGCTGGAATC 540
DB 500 TTAAGGTTTCTGGAAGAACAAACAGCCAAATGCCATATGATGATGCTGGAATC 559
QY 541 CATGCCAGAGAAATGGATCTCTCTGCTGCTTCTGCTTGGTTTCATAGGCCAT----- 591
DB 560 CATGCCAGAGAAATGGATCTCTCTGCTGCTTCTGCTTGGTTTCATAGGCCATATACTCAA 619
QY 592 ----- 591
DB 620 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 679
QY 592 -----AATCAATGTGGAAG 609
DB 680 CCGGTGGTTAATGTGACGGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGGAAG 739
QY 610 AACGTTCTTTCTATGGAACAAATCATTTGATGGAAACAGACCTGGAATAGCAACTTTGTC 669
DB 740 AACGTTCTTTCTATGGAACAAATCATTTGATGGAAACAGACCTGGAATAGCAACTTTGCT 799
QY 670 TCCAACACTGTTGTGAGGAAGGTGCATCCAGTTTCTCATGCTCGGAAACCTACTGTGGA 729

DB 800 TCACAAACTGGTGTGAGGAAGTGTCATCCAGTTCTCATGCTCGGAAACCTACTGTGGA 859
QY 730 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCGAGTGGTAGTTTCTTGAGAAAGAAATATC 789
DB 860 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCGAGTGGTAGTTTCTTGAGAAAGAAATATC 919
QY 790 AACGATTAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATAT 849
DB 920 AACGATTAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATAT 979
QY 850 TCCTATACACGAGTAAGAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 909
DB 980 TCCTATACACGAGTAAGAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA 1039
QY 910 GTTCTGCTATTGACAAAACTAGTAAAAATACAGGTATACACATGGCCATGGCTCAGAA 969
DB 1040 GTTCTGCTATTGAGAAAACTAGTAAAAATACAGGTATACACATGGCCATGGCTCAGAA 1099
QY 970 ACCTTATACCTAGCTCTCTGAGGTGGGAGCATGATGATCTATGATTTGGCATCAATAT 1029
DB 1100 ACCTTATACCTAGCTCTCTGAGGTGGGAGCATGATGATCTATGATTTGGCATCAATAT 1159
QY 1030 TCGTTTACA 1038
DB 1160 TCGTTTACA 1168

RESULT 5
AAT11671
ID AAT11671 standard; DNA; 1749 BP.
XX
AC AAT11671;
XX AC
XX 25-MAR-2003 (updated)
DT 12-APR-1996 (first entry)
XX
XX Human plasma carboxypeptidase B coding sequence.
XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 41..1312
FT CDS /*tag= a
FT /*product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /*tag= b
FT mat_peptide 107..1309
FT /*tag= c
XX
XX US5474901-A.
PN
XX 12-DEC-1995.
XX
XX 19-JUL-1994; 94US-0277540.
XX
XX 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
XX
XX (GETH) GENENTECH INC.
PA
XX Drayna DT, Eaton DL;
XX PI
XX WPI; 1996-039508/04.
DR P-PSDB; AAR30293.
XX
XX Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.

PT haemophilina A
XX PS Disclosure; Figure 4; 40pp; English.
XX CC An antibody which specifically binds human plasma carboxypeptidase B
CC (hPCPB) and does not cross react with other carboxypeptidases is
CC useful for the detection of hPCPB in vitro. The antibody is also
CC used for purifying hPCPB from a sample. Purification comprises
CC passing a sample thought to contain hPCPB over either a column to
CC which antibody has been bound, or a plasminogen affinity column, the
CC eluting the column and then recovering the fraction containing the
CC hPCPB.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX SQ Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Query Match 84.2%; Score 909; DB 17; Length 1749;
Best Local Similarity 89.9%; Pred. No. 2.6e-274;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

QY 1 ATGAAGCTTTGCAGCTTGCAGCTTGTAGCTTGTACCCCAATGTTCTCTCTGAGCAGCATGC 60
DB 41 ATGAAGCTTTGCAGCTTGCAGCTTGTAGCTTGTACCCCAATGTTCTCTCTGAGCAGCATGC 100
QY 61 TTCCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGSCAAGTTCAA 120
DB 101 TTCCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGSCAAGTTCAA 160
QY 121 GTTCTACAGAACTTACTACAAATATGAGATGTTTCTTGGCAGCCGGTAACAGCTGAC 180
DB 161 GTTCTACAGAACTTACTACAAATATGAGATGTTTCTTGGCAGCCGGTAACAGCTGAC 220
QY 181 CTATTGTTGAGAAAGAAACAAAGTCCATTTTGTAAATGCACTGATGTCGCAATGTG 240
DB 221 CTATTGTTGAGAAAGAAACAAAGTCCATTTTGTAAATGCACTGATGTCGCAATGTG 280
QY 241 AAAGCCATTTAAATGTGAGCGGAATCCATGCACTGCTTGTGCGAGAGCTGGAAGAT 300
DB 281 AAAGCCATTTAAATGTGAGCGGAATCCATGCACTGCTTGTGCGAGAGCTGGAAGAT 340
QY 301 CTATTCAACAGCAGATTTTCAACGACACAGTCCAGCCCGAGCTCCCGCATCGTACTAT 360
DB 341 CTATTCAACAGCAGATTTTCAACGACACAGTCCAGCCCGAGCTCCCGCATCGTACTAT 400
QY 361 GACAGATATCACTCACTAAATGAATCTATCTTGGATAGATTTATACTGAGAGCAT 420
DB 401 GACAGATATCACTCACTAAATGAATCTATCTTGGATAGATTTATACTGAGAGCAT 460
QY 421 CCTGATATGCTTCAAAAATCCACATTTGGATTCCTTATGAGAGTACCCACTTATGTT 480
DB 461 CCTGATATGCTTCAAAAATCCACATTTGGATTCCTTATGAGAGTACCCACTTATGTT 520
QY 481 TTAAGGTTTCTGAAAGAAACAAACAGCCAAATAATGCCATATGATTTGACTGTGGAATC 540
DB 521 TTAAGGTTTCTGAAAGAAACAAACAGCCAAATAATGCCATATGATTTGACTGTGGAATC 580
QY 541 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGCCAT----- 591
DB 581 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGCCATATAACTCAA 640
QY 592 ----- 591
DB 641 TTCTATGGGAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 700
QY 592 -----AATCGAATGTGGAGAAAG 609
DB 701 CCGGTGCTTAATGTGGACGGTTATGACTACTCATGAGAAAGAAATCGAATGTGGAGAAAG 760
QY 610 AACGTTCTTTCTATGCGAACAATCATTTGATCGGAACAGACCTGGAATAGCACTTTGTC 669
DB 761 AACGTTCTTTCTATGCGAACAATCATTTGATCGGAACAGACCTGGAATAGCACTTTGCT 820
QY 670 TCCAAACTGCTGTGAGGAAGGTGCATCCAGTTCTCTCGGAAACCTACTGTGGA 729

821 TCCAAACTGCTGTGAGGAAGGTGCATCCAGTTCTCTCGGAAACCTACTGTGGA 880
730 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCACTGGCTAGTTTCTTGAGAAGAAATATC 789
881 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCACTGGCTAGTTTCTTGAGAAGAAATATC 940
790 AACAGATTAAACATACATCAGCATGCTTCACTATCTCCAGCATATAGTGTTCATAT 849
941 AACAGATTAAACATACATCAGCATGCTTCACTATCTCCAGCATATAGTGTTCATAT 1000
850 TCCTATACACGAAGTAAAGCAAAAGCAACCATGAGGAACTGTCTCTAGTAGCAGTGAAGCA 909
1001 TCCTATACACGAAGTAAAGCAAAAGCAACCATGAGGAACTGTCTCTAGTAGCAGTGAAGCA 1060
910 GTTCTGCTTATTCACAAAACCTAGTAAATAACAGGTATACATACATGGCCATGCTCAGAA 969
1061 GTTCTGCTTATTCAGAAAACCTAGTAAATAACAGGTATACATACATGGCCATGCTCAGAA 1120
970 ACCTTATACCTAGCTCTCTGAGGCTGGGAGCATTTGGATCTATGATTTGGGCATCAATAT 1029
1121 ACCTTATACCTAGCTCTCTGAGGCTGGGAGCATTTGGATCTATGATTTGGGCATCAATAT 1180
1030 TCGTTTACA 1038
1181 TCGTTTACA 1189

RESULT 6
AAT62846
ID AAT62846 standard; DNA; 1749 BP.
XX AC AAT62846;
XX DT 25-MAR-2003 (updated)
DT 08-MAY-1997 (first entry)
XX Human plasma carboxypeptidase B coding sequence.
XX Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
XX plasma; plasminogen; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
CDS 41..1312
FT /*tag= a
FT /product= Human PCPB
FT sig_peptide 41..106
FT /*tag= b
FT mat_peptide 107..1309
FT /*tag= c
FT misc_binding 134..177
FT /*tag= d
FT /bound_moeity= 46_bp_probe
XX US5593674-A.
PN 14-JAN-1997.
XX 27-APR-1995; 95US-0430787.
XX 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
PR 27-APR-1995; 95US-0430787.
XX (GETH) GENENTECH INC.
PA Drayna DT, Eaton DL;
PI WPI; 1997-099413/09.
XX DR

P-PSDB; AAW14733.

Using human plasma carboxypeptidase B in blood coagulation - is functionally related to carboxypeptidase A and pancreas carboxypeptidase B

Example 2; Column 37-42; 39pp; English.

This sequence encodes human plasma carboxypeptidase B (PCPB) which has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD. PCPB may be used therapeutically in haemostatic regulation. PCPB is purified from human plasma or by transformed cell culture by extraction using plasminogen bound to a solid phase. (Updated on 25-MAR-2003 to correct PP field.)

Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Query Match 84.2%; Score 909; DB 18; Length 1749;
Best Local Similarity 89.9%; Pred. No. 2.6e-274;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

QY 1 ATGAAGCTTTGCGAGCTTGTGAGTCTTGTACCCATTGTTCTCTCTGTGAGCAGCATGTC 60
DB 41 ATGAAGCTTTGCGAGCTTGTGAGTCTTGTACCCATTGTTCTCTCTGTGAGCAGCATGTC 100
QY 61 TTGCGGTTCCAGAGTGCGCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 120
DB 101 TTGCGGTTTCCAGAGTGCGCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 160
QY 121 GTTCTACAGAACTTACTACAAATATGAGATTGTTCTCTGCGAGCGGTAACAGCTGAC 180
DB 161 GTTCTACAGAACTTACTACAAATATGAGATTGTTCTCTGCGAGCGGTAACAGCTGAC 220
QY 181 CTTATTGTGAAGAAAAAAGTCATTTTTTTGTAATGTCATCTGATGTCGACAAATGTG 240
DB 221 CTTATTGTGAAGAAAAAAGTCATTTTTTTGTAATGTCATCTGATGTCGACAAATGTG 280
QY 241 AAAGCCATTAAATGTGAGCGGAATTCATGCGATGTTCTGTCGAGCAGCGTGAAGAT 300
DB 281 AAAGCCATTAAATGTGAGCGGAATTCATGCGATGTTCTGTCGAGCAGCGTGAAGAT 340
QY 301 CTTATTCAACAGCAGATTTCCACGACACAGTACGCCCCCGAGCTCCGCTCGTACTAT 360
DB 341 CTTATTCAACAGCAGATTTCCACGACACAGTACGCCCCCGAGCTCCGCTCGTACTAT 400
QY 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATACTGAGAGCAT 420
DB 401 GAACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATACTGAGAGCAT 460
QY 421 CTTGATATGCTTACAAAATCCCAATGGATTCCTCATTTGAGAAAGTACCCACTTATGTT 480
DB 461 CTTGATATGCTTACAAAATCCCAATGGATTCCTCATTTGAGAAAGTACCCACTTATGTT 520
QY 481 TTAAGGTTTCTGAAAGAACAAACAGCCAAAATGCCATATGGATTGATCTGGATC 540
DB 521 TTAAGGTTTCTGAAAGAACAAACAGCCAAAATGCCATATGGATTGATCTGGATC 580
QY 541 CATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCAT----- 591
DB 581 CATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATATACTCAA 640
QY 592 ----- 591
DB 641 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 700
QY 592 -----AATCGAATGTGAGAAAG 609
DB 701 CCGGTGGTTAATGTGACGGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGAGAAAG 760
QY 610 AACGGTTCTTCTATGGAACAAATCATTTGTCATCGAACAGACCTGTAATAGCAACTTTGTC 669
DB 761 AACGGTTCTTCTATGGAACAAATCATTTGTCATCGAACAGACCTGTAATAGCAACTTTGTC 820

670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 729
821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGA 880
730 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCGAGTGTGTTCTTTGAGAAGAAATATC 789
881 CTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCGAGTGTGTTCTTTGAGAAGAAATATC 940
790 AACCAAGATTAAAGCATACATCAGCATGTCATTCTATCTCCAGCATATAGTGTTCATAT 849
941 AACCAAGATTAAAGCATACATCAGCATGTCATTCTATCTCCAGCATATAGTGTTCATAT 1000
850 TCCTATACAGCAAGTAAAGCAAAAGACCATGAGAACTGCTCTAGTACGAGTGAAGCA 909
1001 TCCTATACAGCAAGTAAAGCAAAAGACCATGAGAACTGCTCTAGTACGAGTGAAGCA 1060
910 GTTCGTGCTATTGCAAAAACCTAGTAAATAATACAGGTATATACATGCGCCATGGCTCAGAA 969
1061 GTTCGTGCTATTGAGAAAACCTAGTAAATAATACAGGTATATACATGCGCCATGGCTCAGAA 1120
970 ACCTTATACCTAGCTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1029
1121 ACCTTATACCTAGCTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1180
1030 TCGTTTACA 1038
1181 TCGTTTACA 1189

RESULT 7
AAQ41001
ID AAQ41001 standard; cDNA; 1749 BP.
XX AAQ41001;
AC AC
DT 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B gene.
XX PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT sig_peptide 41..106 /*tag= a
FT mat_peptide 107..1312 /*tag= b
FT misc_feature /product= PCPB
FT 134..177 /*tag= c
FT /function= probe
FT /note= "used to obtain full-length clones"
XX
PN USS206161-A.
XX
PD 27-APR-1993.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
XX
PA (GETH) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
XX WPI; 1993-151724/18.
DR P-PSDB; AAR36273.
XX
PT New human plasma carboxypeptidase B - used as haemostatic
PT regulator for clotting blood, partic. for treating blood clotting
PT disorders, e.g. haemophilia
XX

PS Disclosure; Fig 4; 40pp; English.

XX Human plasma carboxypeptidase B was isolated from human plasma and
CC partially sequenced. Oligonucleotide primers were designed based on
CC the partial amino acid sequences. The primers were used in a PCR
CC amplification to identify cDNA encoding PCPB from a human liver cDNA
CC library. The PCR product was capable of encoding the first 37 amino
CC acids of PCPB; a 4mer probe was used to obtain the full-length
CC sequence which, although disclosed in the specification, is not
CC claimed. PCPB inhibits the enzymatic conversion by tPA of
XX plasminogen to plasmin in the presence of fibrinogen.
SQ Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

Query Match 84.08; Score 907.4; DB 14; Length 1749;
Best Local Similarity 89.8; Pred. No. 8.3e-274;
Matches 1032; Conservative 0; Mismatches 6; Indels 111; Gaps 1;

QY 1 ATGAAGCTTTGCGAGCTTGCAGCTTGTGACCTTGTATCCCAATGCTCTCTCTGTGAGCAGCATGTC 60
DB 41 ATGAAGCTTTGCGAGCTTGCAGCTTGTGACCTTGTATCCCAATGCTCTCTCTGTGAGCAGCATGTC 100
QY 61 TTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTTCTCTAGAACCTCTAGGCGAAGTTCAA 120
DB 101 TTCGCGTTCCAGAGTGCCCAAGTTCTAGCTGCTCTTCTCTAGAACCTCTAGGCGAAGTTCAA 160
QY 121 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGCGAGCGCGGTAACAGCTGAC 180
DB 161 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGCGAGCGCGGTAACAGCTGAC 220
QY 181 CTTATTCTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGCAATGTG 240
DB 221 CTTATTCTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGCAATGTG 280
QY 241 AAAGCCCAATTAATGTGAGCGGAATTCATGCACTGCTTGTGCGAGCAGCTGGAAGAT 300
DB 281 AAAGCCCAATTAATGTGAGCGGAATTCATGCACTGCTTGTGCGAGCAGCTGGAAGAT 340
QY 301 CTTATTCAACAGCAGATTTCAACGACACAGTCAGCCCCGAGCTCCGCGATCGTACTAT 360
DB 341 CTTATTCAACAGCAGATTTCAACGACACAGTCAGCCCCGAGCTCCGCGATCGTACTAT 400
QY 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 420
DB 401 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 460
QY 421 CCTGATATGCTTACAAAATCCACATTTGGATCCCTCATTTGAGAAGTACCCACTCTATGTT 480
DB 461 CCTGATATGCTTACAAAATCCACATTTGGATCCCTCATTTGAGAAGTACCCACTCTATGTT 520
QY 481 TTAAGGTTTCTGGAAGAAGAACAAAGCCAAATAATGCCATATGGAATGATCTGTGGAATC 540
DB 521 TTAAGGTTTCTGGAAGAAGAACAAAGCCAAATAATGCCATATGGAATGATCTGTGGAATC 580
QY 541 CATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGTTGTTTATAGGCCAT----- 591
DB 581 CATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGTTGTTTATAGGCCATTAATCAAA 640
QY 592 ----- 591
DB 641 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 700
QY 592 -----AATCGAATGTGAGAAAG 609
DB 701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGAGAAAG 760
QY 610 AACCGTTCTTTCTATGGAACAATCATTTGATCGGAACAGACCTGGAATAGCAACTTTGTC 669
DB 761 AACCGTTCTTTCTATGGAACAATCATTTGATCGGAACAGACCTGGAATAGCAACTTTGTC 820
QY 670 TCCAAACACACTGGTGTGAGGAGGTGCAATCCAGTTTCTCATGCTCGGAACACTACTGTGGA 729
DB 821 TCCAAACACACTGGTGTGAGGAGGTGCAATCCAGTTTCTCATGCTCGGAACACTACTGTGGA 880

QY 730 CTTTATCCTGAGTCAGAACCCAGAGTGAAGGCAAGTGGCTAGTTTCTTTGAGAAGAAATATC 789
DB 881 CTTTATCCTGAGTCAGAACCCAGAGTGAAGGCAAGTGGCTAGTTTCTTTGAGAAGAAATATC 940
QY 790 AACGAGATTAAAGCATACATACATGCAATGCAATCTACTCCAGCATATAGTGTGTTCCATAT 849
DB 941 AACGAGATTAAAGCATACATACATGCAATGCAATCTACTCCAGCATATAGTGTGTTCCATAT 1000
QY 850 TCTATACAGCAAGTAAAGCAAAAGCAACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCA 909
DB 1001 TCTATACAGCAAGTAAAGCAAAAGCAACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCA 1060
QY 910 GTTCGCTGCTATTGACAAAACTAGTAAATAACAGGTATACATACATGSCCATGCTCAGAA 969
DB 1061 GTTCGCTGCTATTGAGAAAACCTAGTAAATAACAGGTATACATGSCCATGCTCAGAA 1120
QY 970 ACCTTATACCTAGCTCTCTGGAGGTGGGACGATTGGATCTATGATTGGGCATCAATAT 1029
DB 1121 ACCTTATACCTAGCTCTCTGGAGGTGGGACGATTGGATCTATGATTGGGCATCAATAT 1180
QY 1030 TCGTTTACA 1038
DB 1181 TCGTTTACA 1189
RESULT 8
AAF18005
ID AAF18005 standard; DNA; 1400 BP.
XX
AC AAF18005;
XX
DT 14-MAR-2001 (first entry)
XX
Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
DE Human; lung cancer associated protein; neuroprotective; cytostatic;
XX KW cardioactive; immunomodulatory; muscular active; vulnery;
XX KW gastrointestinal; nephrotropic; antiinfective; gynecological;
XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX KW proliferative disorder; wound healing; infectious disease; ds.
OS Homo sapiens.
XX
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
DR P-PSDB; AAB58129.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer -
XX
XX Claim 1; Page 507; 1425pp; English.
PS
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the

CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
Query Match 40.8%; Score 441; DB 21; Length 1400;
Best Local Similarity 98.9%; Pred. No. 1.8e-127;
Matches 444; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 590 ATATCGATGTGAGAAACCGTCTTCTATGGAACATCATTCGATCGGACAG 649
DB 177 AGAATCGAATGTGAGAAACCGTCTTCTATGGAACATCATTCGATCGGACAG 236
QY 650 ACCTGAATAGCACTTGTCTCAAAACACATGCTGAGGAAGTGCATCCAGTTCTCAT 709
DB 237 ACCTGAATAGCACTTGTCTCAAAACACATGCTGAGGAAGTGCATCCAGTTCTCAT 296
QY 710 GCTCGAAACCTACTGTGGACTTTATCTGAGTCAGAACAGAGTGAAGCGAGTGGCTA 769
DB 297 GCTCGAAACCTACTGTGGACTTTATCTGAGTCAGAACAGAGTGAAGCGAGTGGCTA 356
QY 770 GTTCTTGAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCTATCTCC 829
DB 357 GTTCTTGAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCTATCTCC 416
QY 830 AGCATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGCATGAGGAAGTGT 889
DB 417 AGCATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGCATGAGGAAGTGT 476
QY 890 CTCTAGTAGCCAGTGAAGCAGTTCGTCTATTGACAAACTAGTAAATAATACCAGGTATA 949
DB 477 CTCTAGTAGCCAGTGAAGCAGTTCGTCTATTGAGAAACTAGTAAATAATACCAGGTATA 536
QY 950 CACATGCCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCT 1009
DB 537 CACATGCCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCT 596
QY 1010 ATGATTGGGCATCAATATTCGTTTACA 1038
DB 597 ATGATTGGGCATCAATATTCGTTTACA 625
RESULT 9
ABX49430
ID ABX49430 standard; cDNA; 416 BP.
XX
AC ABX49430;
DT
XX 21-FEB-2003 (first entry)
DE Bovine EST associated with lactation/muscle/fat deposition #14595.
XX
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
XX Bos Taurus.
OS
XX US2002137139-A1.
PN
XX 26-SEP-2002.
PD
XX 24-SEP-2001; 2001US-0960352.
PF
XX

PR 12-JAN-1999; 99US-115707P.
XX 11-JAN-2000; 2000US-0480902.
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARRE/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX WPI; 2003-110599/10.
DR
XX New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle -
XX
PS Claim 2; SEQ ID No 14595; 245pp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMPD), derived
CC from cattle, and the LMPD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX4947, or complements of them.
CC Also included are: (1) a transformed cell having a nucleic acid
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMPD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMPD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ Sequence 416 BP; 125 A; 86 C; 93 G; 112 T; 0 other;
Query Match 30.7%; Score 331.2; DB 25; Length 416;
Best Local Similarity 87.3%; Pred. No. 2.7e-93;
Matches 363; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 150 GATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGAAACAAAGTCCATTT 209
DB 1 GATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGAAACAAAGTCCATTT 60
QY 210 TTTTGTAAATGCATCTGATGTCACAAATGTGAAGGCCATTTAAATGTGAGCGGAATTC 269
DB 61 TTTTGTAAATGCATCTGATGTAAGCAATGTGAAGGCCATTTAAATGTGAGCGGAATTC 120
QY 270 ATGCACTGCTCTGCTGGCAGACGTGGAAGATCTTTTCAACAGCAGATTTTCAACGACAC 329
DB 121 ATTACGGGCTCTGGTGAAGAAATGTGAAGATCTTTATCCGGCAGCAGACTTCCCAATGACAC 180
QY 330 AGTCAGCCCGGAGCTCCGATCTGATGATGACAGATGATCACTCACTAAATGAATCTA 389
DB 181 CATCAGCCCCGGGATCTCTCTCTACTATGAACAGATGATCACTCACTAAATGAATCTA 240
QY 390 TTCTTGATAGAAATTTTATACTGAGAGGATCTCTGATATGCTTACAAAATTCACATTTGG 449
DB 241 TTCTTGATAGAAATTTTATACTGAGAGGATCTCTGATATGCTTGAAGAAATTCACATTTGG 300
QY 450 ATCCTCATTTGAGAAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAGAACAAACAGC 509


```
FT FT /tag= az
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(32642,T)
FT FT /tag= ba
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(32793..32794,TAG)
FT FT /tag= bb
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT /tag= bc
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(34721,T)
FT FT /tag= bd
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT /tag= be
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(35425,C)
FT FT /tag= bf
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(36050,G)

Query Match 18.9%; Score 204; DB 25; Length 55827;
Best Local Similarity 100.0%; Pred. No. 3.1e-52;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAGGTCATCCAGTCTCTCATGCTCGGAACCTACTGTGACCTTATCTGTAGTCA 744
Db |||||||
Db 42043 GAGGAGGTCATCCAGTCTCTCATGCTCGGAACCTACTGTGACCTTATCTGTAGTCA 42102
QY 745 GAACGAGAGTGAAGGAGGCTGGTGTCTTCTGAGAGAAATATCAACGAGATTAAAGCA 804
Db |||||||
Db 42103 GAACGAGAGTGAAGGAGGCTGGTGTCTTCTGAGAGAAATATCAACGAGATTAAAGCA 42162
QY 805 TACATCAGATGATTCATCTCCAGCATATAGTGTTCATATTCCTATACACGAAGT 864
Db |||||||
Db 42163 TACATCAGATGATTCATCTCCAGCATATAGTGTTCATATTCCTATACACGAAGT 42222
QY 865 AAAGCAAGCAAGCAAGCAAGCAAGT 888
Db |||||||
Db 42223 AAAGCAAGCAAGCAAGCAAGCAAGT 42246
```

```
RESULT 12
ABN19133
ID ABN19133 standard; cDNA; 231 BP.
XX AC
XX ABN19133;
XX
XX 24-JUN-2002 (first entry)
XX DE Human ORFX polynucleotide sequence SEQ ID NO:6743.
XX
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis; gene; ss.
XX OS Homo sapiens.
XX
XX W0200192523-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US10836.
XX
XX 30-MAY-2000; 2000US-206132P.
XX 29-AUG-2000; 2000US-228716P.
XX
```

```
PA
XX
XX Shimketa RA, Leach MD;
DR WPI; 2002-106308/14.
XX P-PSDB; ABP03381.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders
XX
XX Disclosure; SEQ ID 6743; 1037pp; English.
XX
XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
XX in the specification)). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX bone degenerative disorders, or periodontal disease, and for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX reperfusion injury in various tissues and conditions resulting from
XX systemic cytokine damage.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 231 BP; 74 A; 58 C; 40 G; 59 T; 0 other;
```

```
Query Match 15.5%; Score 167; DB 24; Length 231;
Best Local Similarity 82.7%; Pred. No. 6.7e-42;
Matches 191; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 223 TCTGATGTCGACATGTAAGCCCATTTAAATGTGAGCGGAATTCATCGTCGCTTG 282
Db |||||||
Db 1 TCTGATGTCGACAGTGTCAAAGCCTATTTAAATGCGAGCAGAAATTCATTTAACGCTCG 60
QY 283 CTGGCAGAGCTGGAAGATCTTATTCAACAGCAGATTTCCAACGACAGTCAGCCCCCGA 342
Db |||||||
Db 61 ATGAACAACGTGGAGGATCTAATTCACAGCAGCGTCGTCATGTTAGCCCCCGA 120
QY 343 GCCTCCGCGCTGCTACTATGACAGATATCTACTAAATGAAATCTATTCTTGGATAGAA 402
Db |||||||
Db 121 GCCTCTCTCTCATCTACTATGACAGATATCTACTGCTTAAATGAAATCTATTCTTGGATAGAA 180
QY 403 TTTATTAACGTGAGAGGATCTGATGCTTACAAAATCCCAATTCGATGCC 453
Db |||||||
Db 181 GTTATAACTGAACAGCAGCCCTGACATGCTCCAGAAAATCTACATTTGATGCC 231

RESULT 13
ABK62561/c
ID ABK62561 standard; cDNA; 431 BP.
XX
XX ABK62561;
XX
XX 18-JUN-2002 (first entry)
XX
XX Rat sequence differentially expressed in response to a hepatotoxin #468.
XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
```

differential expression; centrilobular necrosis; steatosis.

Rattus norvegicus.

W0200210453-A2.

07-FEB-2002.

30-JUL-2001; 2001WO-US23872.

31-JUL-2000; 2000US-222040P.

02-NOV-2000; 2000US-244880P.

11-MAY-2001; 2001US-290029P.

15-MAY-2001; 2001US-290645P.

22-MAY-2001; 2001US-292336P.

06-JUN-2001; 2001US-295798P.

13-JUN-2001; 2001US-297457P.

19-JUN-2001; 2001US-298884P.

09-JUL-2001; 2001US-303459P.

(GENE-) GENE LOGIC INC.

Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

WPI; 2002-241625/29.

Predicting toxic effects of compounds or the progression of these toxic

effects by determining the changes in gene expression in tissues or

cells exposed to the toxin and comparing these to gene expression in

unexposed tissues or cells -

Claim 1; Seq ID No 468; 239pp; English.

The invention relates to methods for predicting toxic effects of compounds or the progression of these toxic effects by determining the global changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells. Also included are methods of predicting at least one toxic effect of a compound or progression of a toxic effect, preferably the hepatotoxicity of a compound, comprising detecting the level of expression in a tissue or cell sample exposed to the compound of two or more genes listed in the specification, where differential expression of the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilise a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridises to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound or agent. Hepatotoxicity is characterised by centrilobular necrosis and steatosis. The present sequence is an expressed sequence tag (EST) or cDNA derived from a gene which is differentially expressed in response to a hepatotoxic agent.

Seq Sequence 431 BP; 130 A; 85 C; 84 G; 131 T; 1 other;

Query Match 12.0%; Score 129.6; DB 24; Length 431;

Best Local Similarity 87.6%; Pred. No. 5.1e-30;

Matches 141; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 877 CATGAGGAACGTCTCTAGTAGCCAGTGAAGCAGTTCGTCTATTGTGACAAACTAGTAA 936

Db 431 CATGAGGAACGTCTTNTAGTGGCCAGCGAAGCAGTTCGTGCGCATTTGAAAGTATTAAATAA 372

QY 937 AATACAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGAGGTGG 996
|||
Db 371 AACACAGGTATACACATGGCTCAGTGGCTCAGAAAGTTTATATCTAGTCTCTGAGGTCT 312
|||
QY 997 GACGATTGGATCTATGATTTTGGGCATCAAAATATTTCGTTTAC 1037
|||
Db 311 GATGATTGGATCTATGATTTTGGGCATCAAAATATTTCGTTTAC 271
|||

RESULT 14

AAT35760

ID AAT35760 standard; cDNA; 927 BP.

XX AAT35760;

XX 09-NOV-1996 (first entry)

XX Rat mature carboxypeptidase B cDNA.

XX Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.

XX Rattus sp.

XX Key Location/Qualifiers

FT primer_bind complement (1..50)

FT /*tag= a

FT /*note= "Mature CPB 5'-end primer"

FT primer_bind 898..927

FT /*tag= b

FT /*note= "CPB 3' end primer"

XX WO9623064-A1.

XX 01-AUG-1996.

XX 25-JAN-1996; 96WO-US00995.

XX 25-JAN-1995; 95US-0378233.

XX (BIOT-) BIOTECHNOLOGY GEN CORP.

XX Fulga N, Gorecki M, Hartman J, Mendelovitch S;

XX WPI; 1996-362688/36.

XX P-PSDB; AAW00602.

XX Purified active recombinant carboxypeptidase B prodn. - by

expressing DNA encoding the pro-enzyme, folding and enzymatic

cleavage to give active form, used e.g. for insulin prodn.

XX Example 1; Page 37-38; 49pp; English.

XX A cDNA sequence (AAT35760) codes for the mature form (AAW00602) of rat carboxypeptidase B (CPB). A full-length sequence including rat cDNA (AAT35759) coding for the activation peptide (AAW00601) can be obt'd. by PCR amplification (see also AAT35756 and AAT35758) of cDNA derived from Sprague-Dawley rat pancreas. Pro-CPB can be expressed in host cells, e.g. Escherichia coli, and subsequently recovered, CC refolded and cleaved with trypsin to yield the active enzyme. CPB CC produced this way is cheaper than porcine pancreatic enzyme, and is CC free of other proteases.

XX Seq Sequence 927 BP; 253 A; 226 C; 213 G; 235 T; 0 other;

Query Match 11.5%; Score 124; DB 17; Length 927;

Best Local Similarity 57.9%; Pred. No. 4.4e-28;

Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

QY 596 GAATGTGGAGAAAGACCGTCTCTTCTATCGGAACATCTTCATCGGACGACCTGA 655

Db 362 GAATGTGGAGAAAGACCGCTCTACTATGCTGGAAGTTCTCTGTTGGGTAGACCCCA 421

Qy 656 ATAGCAACTTGTCTCAAAACACTGGTGTGAGGAAGTGATCCAGTTCCTCATGTCTCGG 715
Db |||||
422 ACAGGAATTTTAATGC---TGGCTGGTGTGAAGTGGGAGTTCCTCGAGTCCCTGTCTCG 478
Qy 716 AAACCTACTGGGACTTATCTCTAGTCAGACAGCAAGTGAAGGCAAGTGTCTGTCTCT 775
Db |||||
479 AAACCTTACTGTGGACCAAGCCAGAGTCTGAAAAGAGACAAAGGCGCTGGCAGATTCA 538
Qy 776 TGAGAAGAAATATCAACAGATTAAAGCATACATCAGCATGCATTCATCTATCCAGCATA 835
Db |||||
539 TCCGCAACACCTCTCACCATCAAGGCTACCTGACCATCCACTCATCTACAGATGA 598
Qy 836 TAGTGTTCCTATTCCTATACAGCAAGTAAAGCAAGACCATGAGGAACTGTCTCTAG 895
Db |||||
599 TGCTCTACCTTACTCTCTATGACTACAACTGCGCTGAGAACTATGAGGAATTGAATGCC 658
Qy 896 TAGCCAGTGAAGCAAGTTCGTGCTATTGACAAAACTAGTAAANATACAGGTATACACATG 955
Db |||||
659 TGGTGAAGGTGCGGCAAGAGCTTGCC---ACTCTGCATGGCACCAGTACACATATG 715
Qy 956 GCCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGATTGGATCTATGATT 1015
Db |||||
716 GCCAGGAGCTACAACTATCTCTCTGCTGGGGATCTGAGACTGTCTTTATGATC 775
Qy 1016 TGGCATCAAAATATTCGTTTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTC 1075
Db |||||
776 AGGAATCAATATTCCTTTACCTTTGA-ACTCGGGATACAGGCTTCTTTGGCTTCTC 834

RESULT 15

AAQ90601
ID AAQ90601 standard; DNA; 921 BP.

AC AAQ90601;

XX 13-MAR-1996 (first entry)

XX Porcine carboxypeptidase B coding sequence.

KW Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
KW human serum albumin; premating factor alpha; mating factor alpha;
KW proCBP; ds.

XX Sus scrofa.

XX Key Location/Qualifiers
FT 1..921
FT /*tag= a
FT /product= Porcine carboxypeptidase B.

XX W09514096-A1.

XX 26-MAY-1995.

XX 16-NOV-1994; 94WO-US13142.

XX 16-NOV-1993; 93US-0153258.

XX (ELIL) LILLY & CO ELI.

XX Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
PI Sterner JL, Zhang H;

XX WPI; 1995-200386/26.
XX P-PSDB; AAR/5132.

XX DNA encoding porcine carboxypeptidase B - used for transforming
PT host cells, partic. Pichia species, for prodn. of the enzyme

XX Claim 1; Page 20; 34pp; English.

XX The porcine carboxypeptidase B coding sequence can be place in a
CC bacterial or pref. Pichia yeast expression vector. The expression

CC vector further comprises the signal peptide of either human serum
CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC (designated pF489 - NRRL B-21028); mating factor alpha (designated
CC pF474 - NRRL B-21032) or the porcine proCBP signal peptide, for
CC (designated pLGD27 - NRRL B-21027). The method can be used for
CC producing large amounts of porcine carboxypeptidase B and when
CC produced in Pichia yeast, the protein does not need solubilisation
CC or folding. The produced enzyme is then used for pref. cleaving
CC basic residues from the carboxy terminus of proteins.
XX
SQ Sequence 921 BP; 270 A; 238 C; 194 G; 219 T; 0 other;

Query Match 11.4%; Score 123.4; DB 16; Length 921;
Best Local Similarity 56.3%; Pred. No. 6.8e-28;
Matches 273; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

Qy 555 GATCTCTCTGCTTCTGCTTGTGGTTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
Db |||||
321 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCG 380
Qy 615 TTCTTTCTATGCGAACAACTATTCGATCGGAACAGACCTGAATAGCAACTTTCTCTCAA 674
Db |||||
381 CTCTACCAATGTGGAACTACCTGCAATGGCAGACCCCAACAGAAATTTGATGTGG 440
Qy 675 AACTGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
Db |||||
441 G---TGGTGACAACTGGAGCCTCTACAGACCCCTGGATGAGACTTACTGTGATCTGC 497
Qy 735 TCTGAGTCAAGACCAAGAGTGAAGCGAGTGGCTAGTTTCTTGAGAGAAATATCAACCA 794
Db |||||
498 TGCAGAGTCTGAAAAGAGAGACCAAGCCCTGGCTGATTTTATACGCAACACCTCTCTC 557
Qy 795 GATTAAAGCATACATCAGCATGCATTCATCTACTCCAGCATATAGTGTTCATATTCCTA 854
Db |||||
558 CATCAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCCTATTTCCTA 617
Qy 855 TACACGAAGTAAAGCAAGACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
Db |||||
618 TGATTACAACCTCCCGAGNACAATGCTGATTTGAATTAACCTGGCTAAGGCTGCCGTGAA 677
Qy 915 TGCTATTGACAAACTAGTAAATAATACAGGTATACACATGGCCATGGCTCAGAAACCTT 974
Db |||||
678 AGAACTTG---CTACACTGTATGGCCCAAGTACATACGCGCCAGGAGCTACACAAT 734
Qy 975 ATACCTAGCTCTGGAGGTGGGACGATTTGATGATTTGGCATCAAAATATTCGTT 1034
Db |||||
735 CTATCTCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAAATATTCCTT 794
Qy 1035 TACAT 1039
Db 795 CACCT 799

Search completed: January 1, 2004, 11:08:45

Job time : 325.12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:34:47 ; Search time 559.33 Seconds
(without alignments)
2649.143 Million cell updates/sec

Title: US-09-980-881a-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLPIVIFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09980881/runat_31122003_135915_26829/app_query.fasta_1.1628
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=Dits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN 1.1 8325 @runat_31122003_135915_26829 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

RESULT 1

ALIGNMENTS

Result No.	Score	Query Match %	Length	DB ID	Description
1	1884	98.6	1546	9	AB011969 Homo sapi
2	1849.5	96.8	1625	6	AR232183 Sequence
3	1832	95.9	1272	6	AR086324 Sequence
4	1832	95.9	1272	6	BD106653 Method of
5	1832	95.9	1728	6	AX409750 Sequence
6	1832	95.9	1728	6	AX706330 Sequence
7	1832	95.9	1728	9	M75106 Human prepr
8	1832	95.9	1749	6	I16100 Sequence 2
9	1832	95.9	1749	6	I33526 Sequence 2
10	1827	95.6	1272	9	BT006936 Homo sapi
11	1827	95.6	1272	12	BT007990 Synthetic
12	1827	95.6	1715	9	BC007057 Homo sapi
13	1834.5	80.3	1421	10	AB021968 Mus muscu
14	1534.5	80.3	1490	10	AF164524 Mus muscu
15	1528.5	80.0	1430	10	AF186188 Mus muscu
16	1502.5	78.6	1425	10	AB042598 Rattus no
17	700.5	36.7	1327	5	AF190274 Botrhops
18	682	35.7	1441	5	AB09302 Paralicht
19	646	33.8	1251	4	SSC133775 Sus scrof
20	641	33.5	1254	12	BT008111 Synthetic
21	641	33.5	1310	9	HS224866 Homo sapi
22	640.5	33.5	1454	9	BC015338 Homo sapi
23	640.5	33.5	1215	6	I67698 Sequence 1
24	639	33.4	1263	6	A51896 Sequence 60
25	639	33.4	1263	6	A87530 Sequence 11
26	639	33.4	1263	6	AR085861 Sequence
27	631.5	33.0	1284	6	A51913 Sequence 77
28	631.5	33.0	1284	6	AR085875 Sequence
29	628.5	32.9	1442	10	MUSCARA
30	627	32.8	1251	4	DOGZAP47
31	625.5	32.7	1251	9	BC012613 Homo sapi
32	623	32.6	2154	6	A67368 Sequence 12
33	621.5	32.5	1332	6	AX333659 Sequence
34	621.5	32.5	1332	9	HUMPASP
35	621	32.5	1427	10	RNU67914
36	619.5	32.4	1622	9	HUMCARWC
37	619.5	32.4	1622	11	G28614
38	611.5	32.0	1302	6	AX616688 Sequence
39	611.5	32.0	1311	6	AR225025 Sequence
40	611.5	32.0	1311	6	AX106749 Sequence
41	611.5	32.0	1993	6	AX704687 Sequence
42	609.5	31.9	1907	9	BC033684 Homo sapi
43	608.5	31.8	1672	5	BC041315 Xenopus 1
44	604	31.6	927	6	AR072911 Sequence
45	581	30.4	1339	5	GGCBPAMR X64539 G.domesticu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

REFERENCE 1 (bases 1 to 1625)
AUTHORS Gan, W., Ketchum, K.A., Di Francesco, V. and Beasley, E.M.
TITLE Isolated human protease proteins, nucleic acid molecules encoding human protease proteins, and uses thereof
JOURNAL Patent: US 6455294-A 1 24-SEP-2002;
FEATURES Location/Qualifiers
1..1625
source /organism="unknown"
BASE COUNT 488 a 334 c 315 g 488 t
ORIGIN

Alignment Scores:
Pred. No.: 6.52e-173 Length: 1625
Score: 1849.50 Matches: 355
Percent Similarity: 94.18% Conservative: 1
Best Local Similarity: 93.92% Mismatches: 4
Query Match: 96.78% Indels: 18
DB: 6 Gaps: 1

US-09-980-881A-2 (1-360) x AR232183 (1-1625)

Qy 1 MetLysLeuCySerLeuAlaValLeuValProLeuPheCysGluGlnHisVal 20
Db 17 ATGAAGCTTTGCAGCCTTGCAAGTCTAGTCTCTCTAGAACCTCTAGCGAAGTTCAA 136

Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
Db 77 TTCGCGTTTCAGAGTGCCAAAGTCTAGTCTCTCTAGAACCTCTAGCGAAGTTCAA 136

Qy 41 ValLeuGlnAsnLeuThrThrThrGluValLeuLeuProGlnProValThrAlaAsp 60
Db 137 GTTCTCAGAGATCTTACTACAAATATGAGATTTCTCTGGCAGCGGTAACAGCTGAC 196

Qy 61 LeuileValLysLysGlnValHisPheValAsnAlaSerAspValAspVal 80
Db 197 CTTATTGTGAAGAAAACCAAGTCCATTTTGTGAATGCACTCTGATGCGCAATGTG 256

Qy 81 LysAlaHisLeuAsnValSerGlyLeuProCySerValLeuLeuAlaAspValGluAsp 100
Db 257 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTCTCTGCTGGCAGATGTGGAAGAT 316

Qy 101 LeuileGlnGlnGlnLeuSerAspThrValSerProArgAlaSerAlaSerTyr 120
Db 317 CTTATTCAACAGCAGATTTCCACGACACAGTACGCCCCGAGCCTCCGATCGTACTAT 376

Qy 121 GluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHis 140
Db 377 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGCAT 436

Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 437 CCTGATATGCTTACAAAAATCCCAATTTGGATCTCTCAATTTGAGAAAGTACCCACTCTATGTT 496

Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyLe 180
Db 497 TTAAGGTGTTCTGAAAAGAACCAAGACGCCAAAATGCCATATGATGATGCTGGGAATC 556

Qy 181 HisAlaArgGluTyrIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
Db 557 CATGCCAGAGATGATCTCTCTGCTCTCTGCTTGTGTTGTTATAGGCCATATATCGAATG 616

Qy 201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220
Db 617 TGGAGAAAGAACCGTCTCTTCTATGCGCAACAATCATTTGATCGGAACAGACCTGAATAGG 676

Qy 221 AsnPheValSerLysHisTrpCysGluGluClyAlaSerSerSerCysSerGluThr 240
Db 677 AACTTTGCTTCCAAACACTGTGTGTGAGGAAGTGCATCCAGTTCTCTCATGCTCGAAACC 736

Qy 241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
Db 737 TACTGTGGACTTTATCTCTGTAGTCAGAACCAAGAGTGAAGCGGCTAGTTCTTCTTGAGA 796

261 ArgAsnIleLeuGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisVal 280
797 AGAAATATCAACAGATTAAGCATATACATGAGCATGATTCATTCATATCCCATATAGTG 856

Qy 281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
Db 857 TTTCCATATTCCTATATACACGAGTAAAGCAAGACCATGAGGAACTGTCTCTAGTAGCC 916

Qy 301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320
Db 917 AGTGAAGCAGTTCGTCTATTGAGAAAATAGTAAAAATACACAGGTATACATATGCGCAT 976

Qy 321 GlySerGlnThrLeuTyrLeuAlaProGlyGlyGlyAspThrTrpIleTyrAspLeuGly 340
Db 977 GGCTCAGAAAACCTTATACCTAGTCTCTGGAGGTGGGACCATGATGATATGATTTGGGC 1036

Qy 341 IleLysTyrSerPhe 345
Db 1037 ATCAATATTCGTTTACAAATTGAACCTTCGAGATACGGGCACATACGATTTCTGCTGCCG 1096

Qy 346 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1097 GAGCGTTTACATCAAAACCCACTGTAGAGAGCTTTTCCGCTGTCTCTAAAA 1148

RESULT 3
AR086324
LOCUS Sequence 1 from patent US 5985562.
DEFINITION AR086324
ACCESSION AR086324.1 GI:10013090
VERSION AR086324.1
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Morser, M. John. and Nagashima, M.
TITLE Method of detecting thrombotic disease risk associated with plasma carboxypeptidase B polymorphisms
JOURNAL Patent: US 5985562-A 1 16-NOV-1999;
FEATURES Location/Qualifiers
1..1272
source /organism="unknown"
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN

Alignment Scores:
Pred. No.: 2.53e-171 Length: 1272
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-2 (1-360) x AR086324 (1-1272)

Qy 1 MetLysLeuCySerLeuAlaValLeuValProLeuPheCysGluGlnHisVal 20
Db 1 ATGAAGCTTTGCAGCCTTGCAAGTCTAGTCTCTCTAGAACCTCTAGCGAAGTTCAA 60

Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
Db 61 TTCGCGTTTCAGAGTGCCAAAGTCTAGTCTCTCTAGAACCTCTAGCGAAGTTCAA 120

Qy 41 ValLeuGlnAsnLeuThrThrThrGluValLeuTrpGlnProValThrAlaAsp 60
Db 121 GTTCTCAGAGATCTTACTACAAATATGAGATTTCTCTGGCAGCGGTAACAGCTGAC 180

Qy 61 LeuileValLysLysGlnValHisPheValAsnAlaSerAspValAspVal 80
Db 181 CTTATTGTGAAGAAAACCAAGTCCATTTTGTGAATGCACTCTGATGCGCAATGTG 240

Qy 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Db 241 AAAGCCCATTTAAATGTGAGCGGAATTCATCGAGTGTCTTGCTGGCAGAGCTGGAAGAT 300
Qy 101 LeuileGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 301 CTTATTCAACAGCAGATTTCCACGACACAGTCCGAGCGAGCTCCGATCGTACTAT 360
Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPileGluPheIleThrGluArgHis 140
Db 361 GAACAGTATCACTACATAATGAATCTATCTTGGATAGATTTATTAACGTGAGCGAT 420
Qy 141 ProAspMetLeuThrIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 421 CCTGATATGCTTACAAAATCCACATTCGATTCCTCAATTCGAGAAGTACCCACTCTATGT 480
Qy 161 LeuLysValSerGlyLysGluInThrAlaLysAsnAlaIleTyrPileAspCysGlyLe 180
Db 481 TTAAGGTTTCTGGAAGAAGCAACACAGCAAAATGCCATATGATGATGCTGGAAATC 540
Qy 181 HisAlaArgGluTyrPileSerProAlaPheCysLeuTyrPheIleGlyHis----- 197
Db 541 CATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATATAACTCAA 600
Qy 197 ----- 197
Db 601 TTCTATGGGATAATAGGCGCAATATACCAATCTCTCGAGCTTGTGGATTTCTATGTATG 660
Qy 198 -----AsnArgMetTyrArgLys 203
Db 661 CCGGTGTTAATGTGGACGGTTATGACTACTCATGGAAGAAAGATCGAATGTGGAGAAG 720
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 721 AACGTTCTTCTATGCGAACATCAATGTCATCGAACACACCTGAATAGAACTTTGCT 780
Qy 224 SerLysHisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 243
Db 781 TCCAAACACTGGTGTAGGAAGGTGCATCCAGTTCTCTCATGCTCGGAACCTACTGTGGA 840
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
Db 841 CTTTATCTCAGTCAGAACAGAGAGTGAAGCGAGTGGCTAGTTCTTGAAGAAGAAATATC 900
Qy 264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 901 AACAGATTAAAGATACATCAGATCGATCATCTATCCAGCATATAGTTTCCATAT 960
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 961 TCCTATACAGAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCA 1020
Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1021 GTTCGTGCTATTGAGAAAACCTAGTAAAAATACCAAGGTATACATACATGGCCATGGCTCAGAA 1080
Qy 324 ThrLeuTyrLeuAlaProGlyGlyLysAspTyrPileTyrAspLeuGlyLysTyr 343
Db 1081 ACCTTATACCTAGCTCTCGAGGTGGGACGATTTGATCTATGATTTGGGCATCAAAAT 1140
Qy 344 SerPhe-----Th 346
Db 1141 TCGTTTCAATTGAATTCGAGATACGGGCACATACGGAATCTTGCTGCGGAGCGTTAC 1200
Qy 346 rSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1201 ATCAAAACCCACTGTAGAGAGCTTTTGGCGCTGTCTCTAAAA 1243
RESULT 4
LOCUS BD10653 1272 bp DNA linear PAT 18-SEP-2002
DEFINITION Method of detecting thrombotic disease risk.
ACCESSION BD10653
VERSION BD10653.1 GI:23201471
KEYWORDS JP 2002502253-A/1.

synthetic construct
synthetic construct
artificial sequences.
1 (bases 1 to 1272)
Morsers, M.J. and Nagashima, M.
Method of detecting thrombotic disease risk
Patent: JP 2002502253-A 1 22-JAN-2002;
SCHERING AG
PN JP 2002502253-A/1
PD 22-JAN-2002
PF 02-JUN-1998 JP 1999501465
PI 03-JUN-1997 US 08/869057
PR MICHAEL JOHN MORSER, MARIKO NAGASHIMA
PC C1201/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1. 1272
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN
Alignment Scores:
Pred. No.: 2,53e-171 Length: 1272
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 6 Gaps: 2
US-09-980-881A-2 (1-360) x BD10653 (1-1272)
Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 1 ATGAGCTTTGACGCTTCAGTCCCTTGAGTCCCTTGATCCCATTTGCTCTCTGAGCAGCATGTC 60
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 61 TTCGCGTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Qy 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAsp 60
Db 121 GTTCTACAGATCTTACTCAACATATGAGATTTCTCTGGCAGCCGGTAACAGCTGAC 180
Qy 61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Db 181 CTTATTGTGAAGAAAACAAAGTCCATTTTGTAAATGATCTGATGTCGACAAATGTG 240
Qy 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Db 241 AAAGCCCATTTAAATGTGAGCGGAATTCATCGAGTGTCTTGCTGGCAGAGCTGGAAGAT 300
Qy 101 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 301 CTTATTCAACAGCAGATTTCCACGACACAGTCCGAGCGAGCTCCGATCGTACTAT 360
Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPileGluPheIleThrGluArgHis 140
Db 361 GAACAGTATCACTACATAATGAATCTATCTTGGATAGATTTATTAACGTGAGCGAT 420
Qy 141 ProAspMetLeuThrIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 421 CCTGATATGCTTACAAAATCCACATTCGATTCCTCAATTCGAGAAGTACCCACTCTATGT 480
Qy 161 LeuLysValSerGlyLysGluInThrAlaLysAsnAlaIleTyrPileAspCysGlyLe 180
Db 481 TTAAGGTTTCTGGAAGAAGCAACACAGCAAAATGCCATATGATGATGCTGGAAATC 540
Qy 181 HisAlaArgGluTyrPileSerProAlaPheCysLeuTyrPheIleGlyHis----- 197

```
Db 541 CATGCCAGAGATGGATCTCTCCTGCTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600
Qy 197 -----
Db 601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660
Qy 198 -----
Db 661 CCGGTGTTAATGTGGACGGTTATGACTATCTATGGAAGAAATCGAATGTGGAGAAAG 720
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 721 AACCGTCTTCTATGCGACAAATCATTTGCATCGAAGACAGACCTGAATAGAACTTTGCT 780
Qy 224 SerIlyHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 243
Db 781 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGA 840
Qy 244 LeuTyrProGluSerGluProGluValIysAlaValAlaSerPheLeuArgArgAsnIle 263
Db 841 CTTTATCCTGAGTCAGAACCCAGAAAGTGAAGCAGTGGCTAGTTCTTTGAGAGAAATATC 900
Qy 264 AsnGlnIleIysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 901 AACAGATTAAAGCATATACATCAGCATGCATTCATCTCCAGCATATAGTTTCCATAT 960
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 961 TCCTATACAGAACTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCA 1020
Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1021 GTTCGCTGCTATTGAGAAACTAGTAAATAATACAGGATATACACATGGCCATGGCTCAGAA 1080
Qy 324 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleIysTyr 343
Db 1081 ACCTTATACCTAGTCTCTGGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1140
Qy 344 SerPhe-----Th 346
Db 1141 TCGTTTACAATTGAATTCGAGATACGGGCACATACGATTCCTGCTCCCGGACGGTTAC 1200
Qy 346 rSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1201 ATCAAAACCCACTGTAGAGAGCTTTTGCCGCTGCTCTTAAAA 1243

RESULT 5
AX409750
LOCUS AX409750 1728 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2397 from Patent WO0229103.
ACCESSION AX409750
VERSION AX409750.1 GI:21442455
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2397 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source
1. 1728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. M75106"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN
3. 79e-171 Length: 1728
Alignment Scores:
Pred. No.:
```

```
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-2 (1-360) x AX409750 (1-1728)
Qy 1 MetIysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 20 ATGAAGCTTTGCAGCCTTCAGTCTCTGTACCCATTGTTCTCTCTGTGAGCAGCATGTC 79
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 80 TTCGCGTTTCAGAGTGCCCAAGTTCTAGCTGCTCTCTAGAACCTCTAGCAAGTTCAA 139
Qy 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAsp 60
Db 140 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAAAGCTGAC 199
Qy 61 LeuIleValIysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
Db 200 CTTATTGTGAAGAAAAACAAGTCCATTTTGTGTAATGTCATCTGATGTCGACAATGTG 259
Qy 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Db 260 AAAAGCCATTAAATGTGAGCGGAATTCATGTCAGTGTCTTGTGGCAGAGCTGGAAGAT 319
Qy 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 320 CTTATTACAGCAGATTTCCACGACACAGTACGCCCGCCGCTCCGCAATCGTACTAT 379
Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
Db 380 GAACAGTATCACTCACTAATGAAATCTATTCTTGGATAGATTATTAACCTGAGAGCAT 439
Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 440 CTTGATATGCTTACAAAATCCACATTTGGATTCCTCATTTGAGAAAGTACCCACTTATGTT 499
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaIysAsnAlaIleTrpIleAspCysGlyIle 180
Db 500 TTAAGGTTTCTGGAAAAGAACCAACAGCCAAAATGCCATATGGATTGACTGTGGAATC 559
Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 560 CATGCCAGAGATGGATCTCTCCTGCTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619
Qy 197 -----
Db 620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679
Qy 198 -----AsnArgMetTrpArgLys 203
Db 680 CCGGTGTTAATGTGGACGGTTATGACTACTCATATGGAAGAAATCGAATGTGGAGAAAG 739
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 740 AACCGTCTTCTATGCGACAAATCATTTGCATCGAAGACAGACCTGAATAGAACTTTGCT 799
Qy 224 SerIlyHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 243
Db 800 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGA 859
Qy 244 LeuTyrProGluSerGluProGluValIysAlaValAlaSerPheLeuArgArgAsnIle 263
Db 860 CTTTATCCTGAGTCAGAACCCAGAAAGTGAAGCAGTGGCTAGTTCTTTGAGAGAAATATC 919
Qy 264 AsnGlnIleIysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 920 AACAGATTAAAGCATATACATCAGCATGCATTCATCTCCAGCATATAGTTTCCATAT 979
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluLysSerLeuValAlaSerGluAla 303
```

Db	980	TCCTATTACAGAGATAAAGCAACATGAGGA	CTGTCTCTAGTAGCCAGTGAAGCA	1039
Qy	304	ValargAlaIleAepLysTherSerLysAasnThrArgTyrThrHisGlyHisGlySerGlu	323	
Db	1040	GTTCGTGCTATTTCAGAAAACTAGTAAATAATACCAGGTATACACATGCGCATGGCTCGAA	1099	
Qy	324	ThrLeuTyrLeuAlaProGlyGlyGlyAsnAspTrrpileTyrAepLeuGlyLleLysTyr	343	
Db	1100	ACCTTATACCTAGCTCTCGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT	1159	
Qy	344	SerPhe-----	Th 346	
Db	1160	TCGTTTACAATGAACTTCGAGATACCGGCACATACGGATCTCTTGTCGCGGACGCTTAC	1219	
Qy	346	rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys	360	
Db	1220	ATCAAAACCCACCTGTAGAGAGCTTTTTCGCGTCTCTCTAAAA	1262	

RESULT 6

AX706330	LOCUS	AX706330	1728 bp	DNA	linear	PAT 04-APR-2003
	DEFINITION	Sequence 1 from Patent WO03014391.				
	ACCESSION	AX706330				
	VERSION	AX706330.1	GI:295562754			
	KEYWORDS	.				
	SOURCE	Homo sapiens (human)				

SOURCE	ORGANISM
Homo sapiens (human)	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE	1
AUTHORS	Morten, J.E. and Jackson, K.N.
TITLE	Method
JOURNAL	Patent: WO 03014391-A 1 20-FEB-2003; Astrazeneca AB (SE)

```

FEATURES
  Location/Qualifiers
    1. .1728
      /organism="Homo sapiens"

```

```

BASE COUNT      518 a   354 c   338 g   518 t
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

US-09-980-881A-2 (1-360) x AX706330 (1-1728)

[illegible]

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1728)
Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
J. Biol. Chem. 266 (32), 21833-21838 (1991)
JOURNAL 92042093
MEDLINE 1939207
PUBMED 1939207
COMMENT Original source text: Homo sapiens liver cDNA to mRNA.
FEATURES
source
1. 1728
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue type="liver"
1. 1728
/gene="pCPB"
20. 1291
/gene="pCPB"
/note="prepro-plasma carboxypeptidase B"
/codon_start=1
/protein_id="AAA60042.1"
/db_xref="GI:189687"
/translation="MKLCSLAVLPIVLVLFCEQHVFAFQSGQVLAALPRTSRQVQLON
LTITIEVLQPVADLVKKQVHFFVNASDVNDVRAHNLNVSIGPISVLLADVEDLI
QQISNDVTPRASAYEQHSLENIYSWIEFITERHPDMLTKIHGSSFEKPYPLV
KMPGQGTAKNAIWDGIIHAREWI SPACFLWFIQHTQFYGIIGQYTNLLRLVDFY
LVVNVVDGYDYSWKNRMWRKNRSFYANNHCIGTDLNRNPFASKHKEEGASSSCSE
TYCGLYPSREPEKAVSAFLRNINOIKAYLSMHSYSQHLVFPVSYTRSKSKOHEELS
LVASEAVRAIEKTSKRYTHGSETLYLAPGGDDWIYDLGHKYSFTIELRDTGY
GFLUPERYIKTCREAFAAVSKIAHWIRNV"
86. 1288
/gene="pCPB"
/product="plasma carboxypeptidase B"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN
Alignment Scores:
Pred. No.: 3,79e-171 Length: 1728
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 9 Gaps: 2
US-09-980-881A-2 (1-360) x HUMPCPBX (1-1728)
Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 20 ATGAAGCTTTTCAGCCTTTCAGCTCTTGTACCCATTTGTTCTCTCTGTGAGCAGCATGTC 79
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
Db 80 TTCGGTTTCAGAGTGCGCAAGTCTTAGCTGCTCTTCTAGAACCTCTAGCAAGTTCAC 139
Qy 41 ValLeuGlnAenLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
Db 140 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGCGACCGGTACAGCTGAC 199
Qy 61 LeuIleValLysLysLysGlnValHisPhePheValAenAlaSerAspValAspVal 80
Db 200 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGTCATCTGATGTCGACAAATGTG 259
Qy 81 LysAlaHisLeuAenValSerGlyIleProCysSerValLeuLeuAlaAspValClnAsp 100
Db 260 AAAGCCCATTTAAATGTAGCGGAATTCATGCGAGTGTCTTGTGCGCAGCGTGGAGAT 319
Qy 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 320 CTTATTACACAGCAGATTTCCACGACACAGTACGCCCGAGCTCGCATCTGACTAT 379
Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Db 380 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTAGAGGCAAT 439
Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 440 CCGTATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTAGTACCCACTATGTT 499
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAenAlaIleTrpIleAspCysGlyIle 180
Db 500 TTTAAGGTTTCTGGAAGAAGAACCAACAGCCAAAAATGCCATATGATGATTTGAGTGGAAATC 559
Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 560 CATGCCAGAGATGATCTCTCTCTGCTTTCTGCTTGGTTTCATAGGCCATATAACTCAA 619
Qy 197 ----- 197
Db 620 TTCTATGGGAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679
Qy 198 -----AsnArgMetTrpArgLys 203
Db 680 CCGTGGTTAATGTGGACGGTTATGACTACTCATGGAAGAAATCGAATGTGAGAAAG 739
Qy 204 AsnArgSerPheTyrAlaAenAenHisCysIleGlyThrAspLeuAenSerAsnPheVal 223
Db 740 AACGGTTCTTCTATGCGAACAAATCATGTGATCGAAACAGACCTGAAATAGGAATTTGCT 799
Qy 224 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 243
Db 800 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGA 859
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
Db 860 CTTTATCCTGAGTCAGAACCCAGAAAGTGAAGCAGTGTAGTTCTTTGAGAAGAAATATC 919
Qy 264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 920 AACAGATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTTCCTCATAT 979
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 980 TCCATACAGCAAGTAAAGCAAGAACCATGAGGAACTGTCTTAGTACCGCATGGAAGCA 1039
Qy 304 ValArgAlaIleAspLysThrSerLysAenThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1040 GTTCGTCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGCGCATGGCTCAGAA 1099
Qy 324 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
Db 1100 ACCTTATACCTAGCTCCTGAGGTGGGACGATTTGGATCTATGATTTTGGGCATCAATAT 1159
Qy 344 SerPhe-----Th 346
Db 1160 TCGTTTCAATGTAACCTTCAGATACCGGCACATACCGATTCCTGTCGCGGACGGTTAC 1219
Qy 346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1220 ATCAAAACCCACTGTAGAGAAGCTTTTGGCGCTGTCTCTTAA 1262
RESULT 8
116100
LOCUS 116100
DEFINITION Sequence 2 from patent US 5474901.
ACCESSION 116100
VERSION 116100.1 GI:1251008
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Antibodies to human carboxypeptidase B and methods of use thereof
JOURNAL Patent: US 5474901-A 2 12-DEC-1995;


```
Db 1141 TCGTTTAACTTGAATTCGAGATACGGGACACATCGGATTTCTTCTCGCGAGGTTAC 1200
Qy 346 rSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1201 ATCAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTATAA 1243
RESULT 11
LOCUS BT007990 1272 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
carboxypeptidase U) mRNA, partial cds.
ACCESSION BT007990
VERSION BT007990.1 GI:30584818
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1272)
AUTHORS Kalline,N., Chen,X., Rolfe,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1272)
AUTHORS Kalline,N., Chen,X., Rolfe,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
source
1..1272
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/notes="vector: pDNR-Dual"
1..>1272
/notes="Mutations: 1271:Stop->Leu"
/codon_start=1
/transl_table=11
/product="Homo sapiens carboxypeptidase B2 (plasma,
carboxypeptidase U)"
/protein_id="AAF3662.1"
/db_xref="GI:30584819"
/translations="MKLCSLAVLVPIVLFCEQHVFAFQSGVLAALPRTSRQVQLQN
LTWYETVLWQPVTDLIKKKQVFPVNASDVNVKALHNVSGIPCSVLLADVDELI
QQQISNDTVSPRASVYEQYHSLNIYSWIEETTERHPDMLTKIHGSFKEYPLVY
LKVSGKEQAANAIWIDCGTHAREWISPAFLWFIGHTOFYGLIGQYITLLRLVDYF
VMPVNVVDGYDSKNNRMWRKRSFYANNHCIGTDLNRRNFASKHWCSEAGSSSCSE
TCGLYPESEPEVKAVASFLRRNIQKAYISMHSYSQHIYFPFYSTRSKSKDHEELS
LVASEAVRAETKSKNTRYTHGSGTLYLAGGGDDWIYDLGIKYSFTIELRTGTY
GFLLPERYIKPTCREAFAAVSKIAHWIRNVL"
BASE COUNT 372 a 268 c 273 g 359 t
ORIGIN
Alignment Scores: 7.91e-171 Length: 1272
Pred. No.: 1827.00 Matches: 356
Score:
```

```
61 TCGGTTTCAGAGTGGCCAAAGTCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Qy 41 ValLeuGlnAsnLeuThrThrTyxGluileValLeuThrGlnProValThrAlaasp 60
Db 121 GTTCTACAGAATCTTACTACACATATAGAGATTGTCTCTGGCAGCCGGTAACAGCTGAC 180
Qy 61 LeuileValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 80
Db 181 CTTATTGTGAAGAAAACAAGTCAATTTTTTTGTAATGCACTGTGATGTCGACAAATGTG 240
Qy 81 LysAlaHisLeuAsnValSerGlyileProCysSerValLeuLeuAlaAspValGluasp 100
Db 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAGTGTCTTGTCTGGCAGAGCTGAAGAT 300
Qy 101 LeuileGlnGlnGlnLeuSerAsnAspThrValSerProArgAlaSerAlaSerTyTyx 120
Db 301 CTTATTCAACAGCAGATTTCACAGACACAGTCAAGCCCGAGCTCCGCACTGCTACTAT 360
Qy 121 GluGlnTyxHisSerLeuAsnGluileTyxSerTrpIleGluPheIleThrGluArgHis 140
Db 361 GAACAGTATCACTCACTAAATGAATCTATTTCTTGGATAGAATTTATAACTGAGAGGAT 420
Qy 141 ProAspMetLeuThrLysIleHisGlyileGlySerSerPheGluLysTyxProLeuTyxVal 160
Db 421 CCTGATATGCTTACAAAATCCCAATGGATCCTCAITTTGAGAGTAGTACCCCACTCTATGT 480
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyile 180
Db 481 TTAAGGTTTCTGGAAGAAGAACAGACGACCAAAATGCCATATGCAATGATGCTGGAATC 540
Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 541 CATGCCAGAGTAATGCAATCTCTCTGCTTTCTGCTTGGTTGATGAGGTTTCAAGGCCATATACTCAA 600
Qy 197 ----- 197
Db 601 TTCTATGGATAATAGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 660
Qy 198 -----AsnArgMetTrpArgLys 203
Db 661 CCGGTGGTTAATGTGGATGTTATGACTACTCATCGGAAAGAAATCGAATGTGAGAAAG 720
Qy 204 AsnArgSerPheTyxAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 721 AACCGTTCTTTCTATGCGAACATCATGTGCATCGGAACAGACCTCGAATAGGAACCTTGCT 780
Qy 224 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyxTyxCysGly 243
Db 781 TCCAAACACTGTGTGAGGAAGTGCATCCAGTTCTCTATGCTCGGAACCTACTGTGGA 840
Qy 244 LeuTyxProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
Db 841 CTTTATCTGAGTCAGAACAGAGAGTGAAGCAGTGGCTAGTTCTTGGAGAAGAAATATC 900
Qy 264 AsnGlnIleLysAlaTyxIleSerMetHisSerTyxSerGlnHisIleValPheProTyx 283
Db 901 AACAGAGATTAAAGCATACATCAGATGATTCATCTACCTCCAGCATATAGTGTTTCCATAT 960
Qy 284 SerTyxThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 961 TCCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCA 1020
Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyxThrHisGlyHisGlySerGlu 323
Db 1021 GTTCGTGCTATTGAGAAAACCTAGTAAAAATACACAGGTATACACATGCGCATGCTCAGAA 1080
Qy 324 ThrLeuTyxLeuAlaProGlyGlyLysAspTrpIleTyxAspLeuGlyIleLysTyx 343
Db 1081 ACCTTATACCTAGTCTCTGGAGGTGGGACGATTTGATGATTTGGGCATCAATAT 1140
Qy 344 SerPhe-----Th 346
|||||
```


Db	961	TCCTATACAGAAAGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCA	1020
Qy	304	ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu	323
Db	1021	GTTCTGTCTATTGAGAAACTAGTAAATAATACCAGGTATACACATGCGCATGGCTCAGAA	1080
Qy	324	ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr	343
Db	1081	ACCTTATACCTAGTCTCTGGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAAT	1140
Qy	344	SerPhe-----	Th 346
Db	1141	TCGTTTACAATTGAATTCGAGATACGGGCACATACCGATTCTTGCTGCCGAGCGTTAC	1200
Qy	346	rSerAsnProValGluLysLeuLeuProLeuSerLeuLys	360
Db	1201	ATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAAA	1243
RESULT 12			
BC007057 1715 bp mRNA linear PRI 12-JUL-2001			
LOCUS Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495			
DEFINITION IMAGE:3934520, mRNA, complete cds.			
ACCESSION BC007057			
VERSION BC007057.1 GI:13937896			
KEYWORDS MGC.			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE 1 (bases 1 to 1715)			
AUTHORS Strausberg, R.			
TITLE Direct Submission			
JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian			
Gene Collection (MGC), Cancer Genomics Office, National Cancer			
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
USA			
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT Contact: MGC help desk			
Email: cgapbs-r@mail.nih.gov			
Tissue Procurement: CLONTECH			
cDNA Library Preparation: CLONTECH Laboratories, Inc.			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)			
DNA Sequencing by: Sequencing Group at the Stanford Human Genome			
Center, Stanford University School of Medicine, Stanford, CA 94305			
Web site: http://www-shgc.stanford.edu			
Contact: (Dickson, Mark) mcd@paxil.stanford.edu			
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,			
R. M.			
Clone distribution: MGC clone distribution information can be found			
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov			
Series: IRAL Plate: 16 Row: f Column: 21			
This clone was selected for full length sequencing because it			
passed the following selection criteria: matched mRNA gi: 4503004.			
FEATURES			
Location/Qualifiers			
1..1715			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="LocusID:1361"			
/db_xref="taxon:9606"			
/clone="MGC:12495 IMAGE:3934520"			
/tissue_type="Skeletal Muscle"			
/clone_lib="NIH MGC_81"			
/lab_host="DH10B"			
/note="Vector: pDNR-LIB"			
25..1296			
/codon_start=1			
/product="carboxypeptidase B2 (plasma)"			
/protein_id="AAH07057.1"			
/db_xref="GI:13937897"			
/translation="MKLCSLAVLVPVILFCEQHVAFQSGQVLAALPRTSRQVOVLQN			
LTTTYEIVLWQPVTDLIVKKQVHFVNASVDVNVKAHLNVSGIPCSVLLADVEDLI			


```

polyA_site      1421  FLPERYIKPTCAEALAAISKIVVHVRNT"
BASE COUNT      418 a  320 c  305 g  378 t
ORIGIN
Alignment Scores:
Pred. No.:      7.33e-142  Length: 1421
Score:          1534.50    Matches: 302
Percent Similarity: 78.80%  Conservative: 25
Best Local Similarity: 72.77% Mismatches: 32
Query Match:    80.30%     Indels: 56
DB:             10        Gaps: 3

US-09-980-881A-2 (1-360) x AB021968 (1-1421)

Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 17 ATGAAGCTTCATGGCTTGGATCTCTGTAGCCATCATCTCTAT---GAGCAGCATGGC 73
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 74 TTCGCTTTCAGAGTGGCGAGGTTTATCTGCTCTTCCAGAACCTCCAGGCAAGTTCAA 133
Qy 41 ValLeuGlnAsnLeuThrThrThrTyxGluIleValLeuTrpGlnProValThrAlaAsp 60
Db 134 CTACTTCAGAACTTACTACAACTGATGAGTGGTCTCTGGCAGCCAGTGCAGCTGAA 193
Qy 61 LeuIleValLysLysGlnValHisPheValAlaAsnValSerAspValAspVal 80
Db 194 TTCATCGAGAGAGAGAGAGTTCACATTTTGTGATGCTGTGATGTCGACAGTGC 253
Qy 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Db 254 AAAGCGCATTTAAATGTGAGCAGAAATCCATTTAAAGCTTCTGATCAACACAGTGGAGAC 313
Qy 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyxTy 120
Db 314 CTAATTGACAGCAGACACTTTCAATGACACGCTGACGCCCGCGCTCCGCTTCATCTAT 373
Qy 121 GluGlnTyxHisSerLeuAsnGluIleTyxSerTrpIleGluPheIleThrGluArgHis 140
Db 374 GAGCAGTATCACTCGCTAATGAATCTATTCTCGATAGAGTGCATCACTGAACAGCAT 433
Qy 141 ProAspMetLeuThrLysIleHisIleGlySerPheGluLysTyxProLeuTyxVal 160
Db 434 CCTGACATGCTCCAGAAATCTACATCGATCATCATTCGAGAGAGTACCCACTTATGTT 493
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Db 494 TTAAGGTCTCAGGAAAGGAAACAAAGAAATCAAAAATGCCATCTGATCGACTGTGGAATC 553
Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 554 CATGCCAGAGTAATGGAATTCACCTGCTTCTGTTGTGTTGTTATAGGCTACGTGACAA 613
Qy 197 ----- 197
Db 614 TTCCATGGGAAGAAATCTGTATACCAGACTTCTGAGGCAGCTGGATTTCTACATCATG 673
Qy 198 ----- 198
Db 674 CCCGTGATGAACGTGGATGGCTATGACTACACGTGAGGAAAGAAATCGAATGTGGGGAAG 733
Qy 204 AsnArgSerPheTyxAlaAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 734 AACCGCTCTGCTCAAGAAACAACCGCTGCGTGGGACACAGACCTCGAAACAGAACTTCGCT 793
Qy 224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyxTyxGly 243
Db 794 TCCAAACACTGGTGTGAGAAAGTGGCTCAAGTTCCTCTCTCTCTGTAACCTACTGTGGA 853
Qy 244 LeuTyxProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263

```

```

Db 854 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGCAGTGGCTGACTTCTTGGAGAAATAATC 913
Qy 264 AsnGlnIleLysAlaTyxIleSerMetHisSerTyxSerGlnHisIleValPheProTyx 283
Db 914 GACCACATTAAAGCTTACATCATGATCATCATCTATCTCCCAACAATACTGTTTCCCTAT 973
Qy 284 SerTyxThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 974 TCCATTAACAGAACAAAGCAGGACCAAGAACTGTCTCTAGTGGCCAGGCAAGCA 1033
Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyxThrHisGlyHisGlySerGlu 323
Db 1034 GTTCGTGCAATTTGAAGTATTATATAAAACACCAAGGTACACACGCGCAGTGGCTCAGAA 1093
Qy 324 ThrLeuTyxLeuAlaProGlyGlyAspAspTrpIleTyxAspLeuGlyIleLysTyx 343
Db 1094 AGTTTATATCTAGCTCTCTGGAGGTTCTGACGATGGATCATGATTTGGGCATCAATAT 1153
Qy 344 SerPhe-----Th 346
Db 1154 TCGTTTACAAATTGAGCTCCGAGATACAGGCAGATACGGATTTCTGCTGCTGAGAGATAC 1213
Qy 346 rSerAsnProValGluLysLeuProLeuSerLeuLys 360
Db 1214 ATCAAAACCACTTGTGAGAAAGCTTTGGCCGCTCTCTAAAA 1256

RESULT 14
LOCUS      AF164524          1490 bp      mRNA      linear      ROD 05-APR-2000
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)
ACCESSION  AF164524
VERSION    AF164524.1
KEYWORDS   GI:7416966
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 1490)
AUTHORS    Marx,P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van
            Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
TITLE      Characterization of mouse thrombin-activatable fibrinolysis
            inhibitor
JOURNAL    Thromb. Haemost. 83 (2), 297-303 (2000)
MEDLINE    20201996
PUBMED     10739389
REFERENCE  2 (bases 1 to 1490)
AUTHORS    Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
TITLE      Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100,
            Utrecht 3584 CX, The Netherlands
JOURNAL    Location/Qualifiers
FEATURES   1..1490
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:477035"
            1..1490
            /gene="Tafi"
            84..1352
            /gene="Tafi"
            /function="carboxypeptidase"
            /codon_start=1
            /product="thrombin-activatable fibrinolysis inhibitor"
            /protein_id="AAF62385.1"
            /db_xref="GI:7416967"
            /translation="MKLHGLILVAILLYHQHGFQSGVLSALPRTSRQVLQML
            TTTVEVLWQPVTAETFEKKVEHFFVNASGVDVSKAHNVSRIPFNVLMMNVDELIE
            QCFNDTVSPRASVYEQYHSLNIEYSLWIEVTEHQHPDMLQKTYIGSSPEKPLVYL
            KVSKEQRINKAIWIDGIAHREWIWISPAFLWIFGYTQFHGKLENLTRLLRHVDVFI
            MPVNVVDGYDYTWKKNRWRNRSHKNNRCVGDLDNRNFASKHWCSEKSSSCSET
            YCGLPESBEPKAVADFLRRNDIHKAYISMHSYQQIILFPYSYNRKSKDKHELSL

```

VASEAVRAIESINKNRYTHGSGESLYLAPGSGDDWIYDLGKYSFTIELDRTRYG
FLPERVTKPTCAALAAISKIVHVRINT"
BASE COUNT 438 a 337 c 322 g 393 t
ORIGIN

Alignment Scores:

Pred. No.: 7,8e-142 Length: 1490
Score: 1534.50 Matches: 302
Percent Similarity: 78.90% Conservative: 25
Best Local Similarity: 72.77% Mismatches: 32
Query Match: 80.30% Indels: 56
DB: 10 Gaps: 3

US-09-980-881A-2 (1-360) x AF164524 (1-1490)

```
Qy 1 MetLysLeuCySerLeuAlaValLeuValProLeValLeuPheCysGluGlnHisVal 20
Db 84 ATGAAGCTTCATGCGCTTGGAAATCCTGGTAGCCATCATCTCTAT--GAGCAGCATGGC 140
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
Db 141 TTGCGCTTTCAGATGGCCAGGTTTATCTCTCTTCCAGAACCTCCAGGCAAGTTCAA 200
Qy 41 ValLeuGlnAenLeuThrThrTyrGluLeValLeuTyrGlnProValThrAlaAsp 60
Db 201 CTACTTCAGATCTTACTACACGATGATGAGTGGTCTCTCGCAGCCAGTCACAGCTGAA 260
Qy 61 LeuLeValLysLysLysGlnValHisPhePheValAenAlaSerAspValAspVal 80
Db 261 TTCATCGAGAGAGAGAGAGAGTCCACTTTTGTGAATGCGTCTGTATGTCGACAGTGC 320
Qy 81 LysAlaHisLeuAenValSerGlyLeProCySerValLeuLeuAlaAspValGluAsp 100
Db 321 AAAGCGCATTTAATGTGAGCAGAAATCCATTTACGGTTCTGTGATGAACAGCTGGAGAC 380
Qy 101 LeuLeGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 381 CTAAATGAACAGCAGACCTTCAATGACACGCTGAGCCCGCGCTCCGCTTCATACTAT 440
Qy 121 GluGlnTyrHisSerLeuAenGluLeuTyrSerTrpIleGluPheLeuThrGluAargHis 140
Db 441 GAGCAGTATCATCTCGCTAAATGAAATCTATTCTCGTAGAGTGAAGTCAATCAACAGCAT 500
Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 501 CCTGACATGCTCCAGAAATCTACATCGATCATCATTCGAGAGATCCCACTTTATGTT 560
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyLe 180
Db 561 TTAAAGGTCTCAGAAAGAGAACAAAGATCAAAATGCCATCTGGATCGACTGTGGAATC 620
Qy 181 HisAlaArgGluTyrTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 621 CATGCCAGAGATGGATTTACCTGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 680
Qy 197 ----- 197
Db 681 TTCCATGGGAAGAAATCTGTATACAGACTTCTGAGGACCGTGGATTTCTATCATCATG 740
Qy 198 ----- 203
Db 741 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAGAAAGATCGAATCTGGAGGAG 800
Qy 204 AsnArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAenSerAsnPheVal 223
Db 801 AACCGCTCTGCTCACAAGAACAAACCGCTGCGTGGGACAGACCTGAAACAGGAACTTCGCT 860
Qy 224 SerLysHisTrpCysGluGluValAserSerSerCysSerGluThrTyrCysGly 243
Db 861 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCTCTCTGCTCTGAAACCTACTGTGA 920
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
```

```
Db 921 CTTTATCTCTGAGTCTGAGCCAGGCTGAGGACGCTGCTTCTTCTGAGAGAAATATC 980
Qy 264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 981 GACCACATTAAAGCTTATCATCAGTATGCTACTTCTCCCAACAATACTGTTTCCCTAT 1040
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 1041 TCCTATAACAGACGAAAGACAGGACCAAGAGACTGTCTCTAGTGGCCAGGAGCA 1100
Qy 304 ValArgAlaLeuAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1101 GTTCGTCAATTGAAGTATTATAATAAAACACACAGGTACACACACGCGAGTGGCTCAGAA 1160
Qy 324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
Db 1161 AGTTTATATCTAGCTCTCTGAGGTTCTGACGATTTGGATCTATGATTTGGGCATCAATAT 1220
Qy 344 SerPhe-----Th 346
Db 1221 TCGTTTACAATTGAGTCCGAGATACAGGACGATACGATTCCTGCTGCGCTGAGAGATAC 1280
Qy 346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1281 ATCAACCCACTTGTGCAAGGCTTTGGCGCCATCTCTATAAA 1323
RESULT 15
AF186188 1430 bp mRNA linear ROD 01-OCT-1999
LOCUS Mus musculus carboxypeptidase U mRNA, complete cds.
DEFINITION AF186188
ACCESSION AF186188.1 GI:6003651
VERSION
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1430)
AUTHORS He, Y.C. and Broze, G.
TITLE Isolation and characterization of mouse liver carboxypeptidase B
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1430)
AUTHORS He, Y.C. and Broze, G.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1999) Division of Hematology, Washington
University School of Medicine, 216 S. Kingshighway, St. Louis, MO
63110, USA
FEATURES
Location/Qualifiers
1..1430
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B6/CBAF1J"
/db_xref="taxon:10090"
20..1288
/notes="pro-carboxypeptidase-B; thrombin-activatable
fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds
plasminogen"
/codon_start=1
/product="carboxypeptidase U"
/protein_id="AA00528.1"
/db_xref="GI:6003652"
/translational="MKLHGLILVALIILYEQHGFQFOSGVLSPALPETSQVOLLQNL
TTTYEVLMQFVTAETFEKKKEVHFFVNASDVSVKALNVSRIPFNVLMNVEDLIE
QOTFNDTVPRASAYEYQHSNLSIWEIYEVTEQHPDMLQKLIYIGSSPEKYPVYL
KVSQEKRIKNAIWDICGIIHAREWISPAFLWFIYQTFQHGKNIYLRLLRHPFYI
MPVNVVDGYDVTWKKNRMKNSAHKNRNCVGTDLNRNPFASKHWCSEKSSCSSET
YCGLYPSEPEVKAVADFLRNIDHIKAYISMHSYSQOILFPYSYNKSKDHLELSL
VASEAVRAIGSPFNKTRTYHSGSESILYAPGSGDDWIYDLGKYSFTIELDRTRYG
FLPERVTKPTCAALAAISKIVHVRINT"
BASE COUNT 421 a 320 c 310 g 379 t
ORIGIN
```

Alignment Scores:

Pred. No.:	2.9e-141	Length:	1430
Score:	1528.50	Matches:	301
Percent Similarity:	78.5%	Conservative:	25
Best Local Similarity:	72.5%	Mismatches:	33
Query Match:	79.9%	Indels:	56
DB:	10	Gaps:	3

US-09-980-881A-2 (1-360) x AF186188 (1-1430)

Qy	1	MetLysIeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal	20
Db	20	ATGAAGCTTTCATGGCTTTGGAACTCTGTGTAGCCATCATCTCTCTAT--GAGCAGCATGGC	76
Qy	21	PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln	40
Db	77	TTCCGCTTTACAGTGGCCAGGTTTATCTGCTCTTCCAAAGAACTCTCAGGCAAGTTCAA	136
Qy	41	ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp	60
Db	137	CTACTTCAGATCTTACTTACAACGTATGAGTCTGTTCTTGGCAGCCAGTGACAGCTGAA	196
Qy	61	LeuIleValLysLysLysGlnValHisPheValAsnAlaSerAspValAsnVal	80
Db	197	TTTCATCGAAGAAAGAAAGTCCACTTTTGTGAAATGGCTCTGATGTCACACAGTGTG	256
Qy	81	LysAlaHisIeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp	100
Db	257	AAAGCGCATTTAAATGTGAGCAGAAATTCATTTAAAGTTCTGATGAAACAACGTGGAGAC	316
Qy	101	LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr	120
Db	317	CTAATTGAACAGCAGACTTTCATGTACACGGTCAAGCCCGCGCTCTCCGCTTCATCTAT	376
Qy	121	GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis	140
Db	377	GAGCAGTATCATCTCGCTAAATGAAATCTATTCTCTGGATAGAAGTCTTAACTGAACAGCAT	436
Qy	141	ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal	160
Db	437	CCTGACATGCTCCAGAAATCTTACATCGGATCATCATTCGAGAGTACCACCTTTATGTT	496
Qy	161	LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle	180
Db	497	TTAAAGGTCTCAGGAAGGAACAAGAAATCAAAATGCCATCTGGATCGACTGTGGAATC	556
Qy	181	HisAlaArgGluTyrIleSerProAlaPheCysLeuTrpPheIleGlyHis	197
Db	557	CATGCCAGAAATGGATTTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA	616
Qy	197	-----	197
Db	617	TTCCATGGGAAGAAAATCTGTATACACAGACTTCTGAGGCACGTGGATTTCTACATCATG	676
Qy	198	-----AsnArgMetTrpArgLys	203
Db	677	CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAAATCGAATGTGGAGGAAG	736
Qy	204	AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal	223
Db	737	AACCGCTCTGTCTCAACAGAACAAACCGCTCGTGGGGCAGACCTGAACAGGAACCTTCGT	796
Qy	224	SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly	243
Db	797	TCCAAACAACTGGTGTGAAAGAGGTGGCTCAAGTTCCTCTGCTCTGAAACCTTACTGTGGA	856
Qy	244	LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle	263
Db	857	CTTTATCTTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAAGAAGAATATC	916
Qy	264	AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr	283
Db	917	GACCACATTAAGCTTACATCAGTATGACATCATACTCCCAACAATACTTTTCCCTAT	976

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:00:56 ; Search time 62.7557 Seconds
(without alignments)
910.539 Million cell updates/sec

Title: US-09-980-881A-2

Perfect score: 1911

Sequence: 1 MKLCSLAVLPVILFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1911	100.0	360	21	AA111457 Human brain carbox
2	1822	95.3	386	24	ABG72095 Human protease pro
3	1817	95.1	386	24	ABG72093 Human protease pro
4	1799.5	94.2	423	17	AAR30293 Human plasma carbo
5	1799.5	94.2	423	18	AAW14733 Human plasma carbo
6	1799.5	94.2	423	20	AAW92270 Human plasma carbo
7	1794.5	93.9	423	14	AAR36273 Human plasma carbo
8	1785	93.4	338	21	AA111458 Human brain carbox
9	1325	69.3	246	21	AA111459 Human brain carbox

10	803.5	42.0	211	21	AA588129 Lung cancer associ
11	641	33.5	417	24	AB899688 Amino acid sequenc
12	640.5	33.5	404	16	AA751131 Porcine Tyr-His-Me
13	639	33.4	415	17	AAW06172 Human pancreatic c
14	639	33.4	415	19	AAW74476 Human pancreatic c
15	631.5	33.0	424	17	AAW06175 ProHCPB with PalB
16	629.5	32.9	324	22	AA66562 Bothrops jararaca c
17	623	32.6	716	18	AAW41414 PreproHCPB-linker-
18	619.5	32.4	417	23	AAU84325 Protein CPA3 diffe
19	618.5	32.4	667	24	ABP97414 Anti-human seminal
20	604	31.6	307	17	AAW00602 Rat carboxypeptida
21	604	31.6	434	24	AB899687 Amino acid sequenc
22	604	31.6	437	22	AA800502 Human carboxypepti
23	604	31.6	437	23	AA819175 Human carboxypepti
24	576	30.1	306	16	AA751132 Human protease, PR
25	576	30.1	306	22	AAU04477 Porcine carboxypep
26	563	29.5	354	22	AA66560 Porcine carboxypep
27	563	29.5	354	22	AA66566 Human secreted met
28	563	29.5	362	22	AA674682 Human secreted met
29	563	29.5	374	22	AA665647 Human protease and
30	563	29.5	374	22	AA665655 Human secreted met
31	563	29.5	374	23	AAU82703 Amino acid sequenc
32	563	29.5	374	24	ABJ37888 NOVX protein sequ
33	563	29.5	374	24	ABG74658 Human zinc-binding
34	563	29.5	444	23	AAE24240 Human 23566 (Carbo
35	561	29.4	329	17	AAW06174 Mature HCPB with P
36	561	29.4	349	17	AAW06173 Mature HCPB (His)6
37	554	29.0	349	17	AAW06181 Modified HCPB (D25
38	554	29.0	349	18	AAW13749 Carboxypeptidase B
39	553	28.9	349	17	AAW06182 Modified HCPB (D25
40	553	28.9	349	18	AAW13750 Carboxypeptidase B
41	550	28.8	349	18	AAW13763 Carboxypeptidase B
42	548	28.7	349	18	AAW13752 Carboxypeptidase B
43	548	28.7	349	18	AAW13758 Carboxypeptidase B
44	547	28.6	349	18	AAW13751 Carboxypeptidase B
45	547	28.6	374	24	ABJ37887 NOVX protein sequ

ALIGNMENTS

RESULT 1

AA111457
ID AA111457 standard; Protein; 360 AA.

XX AA111457;

XX AA111457;

DT 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B protein.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; antialzheimers; neurotropic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma.

OS Homo sapiens.

XX WO2000066717-A1.

XX 09-NOV-2000.

PD 01-MAY-2000; 2000WO-JP02878.

PF 30-APR-1999; 99JP-0125169.

XX (MATS/) MATSUMOTO A.

XX Matsumoto A;

XX WPI; 2000-687534/67.

DR N-PSDB; AAC81962.

XX Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
XX
XX Claim 1; Page 68-71; 84pp; Japanese.
XX
XX This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 360 AA;
Query Match 100.0%; Score 1911; DB 21; Length 360;
Best Local Similarity 100.0%; Pred. No. 9.1e-188;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTYEIVLMQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTYEIVLMQPVTD 60
QY 61 LIVKKQVHFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASASY 120
DB 61 LIVKKQVHFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASASY 120
QY 121 EQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCI 180
DB 121 EQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCI 180
QY 181 HAREWISPAFLWFIHGNRMWRKNRSPYANNHCIGTDLNSNFVSKHWCESGSSSCSET 240
DB 181 HAREWISPAFLWFIHGNRMWRKNRSPYANNHCIGTDLNSNFVSKHWCESGSSSCSET 240
QY 241 YCGLYPSSEPEVKAVASFLRNINQIKAYISMHSYSGHIVPPYSYTRSKSDHEELSLVA 300
DB 241 YCGLYPSSEPEVKAVASFLRNINQIKAYISMHSYSGHIVPPYSYTRSKSDHEELSLVA 300
QY 301 SEAVRAIDKTSKTRYTHGHGSETLYLAPGGDDMIYDLGKYSFTSNPPVEKLLPLSLK 360
DB 301 SEAVRAIDKTSKTRYTHGHGSETLYLAPGGDDMIYDLGKYSFTSNPPVEKLLPLSLK 360
RESULT 2
ABG72095
ID ABG72095 standard; Protein; 386 AA.
XX AC ABG72095;
XX AC
XX AC
DT 14-FEB-2003 (first entry)
XX DE Human protease protein, variation #2.
XX
XX Human; enzyme; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW single nucleotide polymorphism; SNP.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 169
FT /note= "This residue is the minor variant of a single
FT nucleotide polymorphism. The major variant is an
XX Ala residue"
XX
PN US2002137179-A1.

XX 26-SEP-2002.
PD
XX
XX 21-MAR-2001; 2001US-0813133.
PF
XX
XX 21-MAR-2001; 2001US-0813133.
PR
XX
XX (GANW/) GAN W.
PA (KETC/) KETCHUM K A.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Gan W, Ketchum KA, Di Francesco V, Beasley EM;
XX WPI; 2003-102387/09.
DR N-PSDB; ABX13670, ABX13671.
XX
XX New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
PT -
XX
PS Claim 1; Page -; 80pp; English.
XX
XX The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for
CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorders. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC protein, variant #2, the gene for which is located on chromosome 13.
CC Note: This sequence is not shown in the specification but has been
CC created by the indexer from information given in figure 3.
XX
SQ Sequence 386 AA;
Query Match 95.3%; Score 1822; DB 24; Length 386;
Best Local Similarity 98.8%; Pred. No. 1.5e-178;
Matches 342; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTYEIVLMQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTYEIVLMQPVTD 60
QY 61 LIVKKQVHFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASASY 120
DB 61 LIVKKQVHFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASASY 120
QY 121 EQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCI 180
DB 121 EQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCI 180
QY 181 HAREWISPAFLWFIHGNRMWRKNRSPYANNHCIGTDLNSNFVSKHWCESGSSSCSET 240
DB 181 HAREWISPAFLWFIHGNRMWRKNRSPYANNHCIGTDLNSNFVSKHWCESGSSSCSET 240

QY 241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSTRSKSKDHEELSLVA 300
DB 241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSTRSKSKDHEELSLVA 300
QY 301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346
DB 301 SEAVRAIEKISKTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346

RESULT 3
ABG72093
ID ABG72093 standard; Protein; 386 AA.
XX AC ABG72093;
XX 14-FEB-2003 (first entry)
XX Human protease protein, variation #1.
XX Human; enzyme; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW single nucleotide polymorphism; SNP.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Misc-difference 169 /note= "This residue is the major variant of a single
FT nucleotide polymorphism. The minor variant is an
FT Thr residue"

XX US2002137179-A1.
XX 26-SEP-2002.
XX 21-MAR-2001; 2001US-0813133.
XX 21-MAR-2001; 2001US-0813133.
XX (GANW/) GAN W.
PA (KETC/) KETCHUM K A.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX Gan W, Ketchum KA, Di Francesco V, Beasley EM;
XX WPI; 2003-102387/09.
XX N-PSDB; ABX13670, ABX13671.
XX New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
XX
XX Claim 1; Fig 2; 80pp; English.

XX The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for

CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorders. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC protein, variant #1, the gene for which is located on chromosome 13.
XX
SQ Sequence 386 AA;
Query Match 95.1%; Score 1817; DB 24; Length 386;
Best Local Similarity 98.8%; Pred. No. 4.9e-178;
Matches 341; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 MKLCSLAVLPIVLFCQHVFAFQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPTAD 60
DB 1 MKLCSLAVLPIVLFCQHVFAFQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPTAD 60
QY 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASY 120
DB 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASY 120
QY 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGSKDQTAKNAIWIDCGI 180
DB 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGSKDQTAKNAIWIDCGI 180
QY 181 HAREWISPAFLWFIQHNRWNRKRSFYANNHCIGTDLNSNFVSKHWCESGASSSCSET 240
DB 181 HAREWISPAFLWFIQHNRWNRKRSFYANNHCIGTDLNSNFVSKHWCESGASSSCSET 240
QY 241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSTRSKSKDHEELSLVA 300
DB 241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSTRSKSKDHEELSLVA 300
QY 301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346
DB 301 SEAVRAIEKISKTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346

RESULT 4
AAR90293
ID AAR90293 standard; Protein; 423 AA.
XX AC AAR90293;
XX 25-MAR-2003 (updated)
DT 12-APR-1996 (first entry) ✓
XX Human plasma carboxypeptidase B.
XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Active-site 181 /note= "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 183 /note= "Residue suspected of being part of the
FT binding site of enzyme."
FT Binding-site 256..257 /note= "Residues suspected of being part of the
FT binding site of enzyme."
FT Active-site 310 /note= "Residue suspected of being involved in
FT catalytic activity of enzyme."
FT Binding-site 312 /note= "Residue suspected of being part of the
FT binding site of enzyme."

FT Active-site 369 /note= "Residue suspected of being involved in catalytic activity of enzyme."
FT Binding-site 370..371 /note= "Residues suspected of being part of the binding site of enzyme."
FT Binding-site 394 /note= "Residue suspected of being part of the binding site of enzyme."
XX US5474901-A.
XX 12-DEC-1995.
XX 19-JUL-1994; 94US-0277540.
XX 01-FEB-1991; 91US-0649591.
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX (GETH) GENENTECH INC.
XX Drayna DT, Eaton DL;
XX WPI; 1996-039508/04.
XX N-PSDB; AAT11671.
XX Antibody to human plasma carboxypeptidase B - useful for detecting and purifying hPCPB for use in treating clotting disorders e.g. haemophilia A
XX Disclosure; Figure 4; 40pp; English.
XX An antibody which specifically binds human plasma carboxypeptidase B (hPCPB) and does not cross react with other carboxypeptidases is useful for the detection of hPCPB in vitro. The antibody is also used for purifying hPCPB from a sample. Purification comprises passing a sample thought to contain hPCPB over either a column to which antibody has been bound, or a plasminogen affinity column, eluting the column and then recovering the fraction containing the hPCPB.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX Sequence 423 AA;
Query Match 94.2%; Score 1799.5; DB 17; Length 423;
Best Local Similarity 89.6%; Pred. No. 3.6e-176;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
QY 1 MKLCSLAVLPVILFCQHVPFQFQVLAALPRTSRQVQLQTLTTTTEIVLWQPTAD 60
DB 1 MKLCSLAVLPVILFCQHVPFQFQVLAALPRTSRQVQLQTLTTTTEIVLWQPTAD 60
QY 61 LIVKKQVHFVNASDVNDVNKAHLNVSGIPCSVLADVLDLQOISNDTVSPRASAYY 120
DB 61 LIVKKQVHFVNASDVNDVNKAHLNVSGIPCSVLADVLDLQOISNDTVSPRASAYY 120
QY 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCI 180
DB 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAIWDGCI 180
QY 181 HAREWISPAFLWIGH-----NRWRK 203
DB 181 HAREWISPAFLWIGHITOPYGIIGQTNLLRLVDFVMPVNVNVDGQDYSKNRMWRK 240
QY 204 NRSFYANNHCIGTDLNENSVKWCERGASSSCSETYCGLYPSEPEVKAVASFLRNI 263
DB 241 NRSFYANNHCIGTDLNENRFASKWCERGASSSCSETYCGLYPSEPEVKAVASFLRNI 300
QY 264 NQIKAYISMHSYSQHIYFPYSTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
DB 301 NQIKAYISMHSYSQHIYFPYSTRSKSDHEELSLVASEAVRAIETKTSKNTRYTHGHGSE 360

QY 324 TLYLAPGGDDWYLDLGIKYSFT 346
DB 361 TLYLAPGGDDWYLDLGIKYSFT 383
RESULT 5
AAW14733
ID AAW14733 standard; Protein; 423 AA.
XX AAW14733;
XX 25-MAR-2003 (updated)
DT 08-MAY-1997 (first entry)
XX Human plasma carboxypeptidase B.
XX Human; plasma carboxypeptidase B; PCPB; haemostatic regulation; plasma; plasminogen.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Peptide 1..22 /note= "Signal peptide"
FT Protein 23..423 /note= "Mature PCPB"
FT Cleavage-site 114 /note= "Potential clip site for activation of PCPB as a carboxypeptidase"
FT Active-site 181 /note= "Expected to be involved in catalytic activity"
FT Binding-site 183 /note= "Expected to be involved in substrate binding"
FT Active-site 184 /note= "Expected to be involved in catalytic activity"
FT Binding-site 256 /note= "Expected to be involved in substrate binding"
FT Binding-site 257 /note= "Expected to be involved in substrate binding"
FT Active-site 310 /note= "Expected to be involved in catalytic activity"
FT Binding-site 312 /note= "Expected to be involved in substrate binding"
FT Active-site 369 /note= "Expected to be involved in catalytic activity"
FT Binding-site 370 /note= "Expected to be involved in substrate binding, thought to determine specificity of PCPB as a carboxypeptidase B"
FT Binding-site 371 /note= "Expected to be involved in substrate binding"
FT Binding-site 394 /note= "Expected to be involved in substrate binding"
XX US5593674-A.
XX 14-JAN-1997.
XX 27-APR-1995; 95US-0430787.
XX 01-FEB-1991; 91US-0649591.
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX 27-APR-1995; 95US-0430787.
XX (GETH) GENENTECH INC.
XX Drayna DT, Eaton DL;
XX WPI; 1997-099413/09.
XX N-PSDB; AAT62846.

XX Using human plasma carboxypeptidase B in blood coagulation - is
PT functionally related to carboxypeptidase A and pancreas
PT carboxypeptidase B
XX
XX
XX Example 2; Column 37-42; 39pp; English.
XX
XX This sequence represents human plasma carboxypeptidase B (PCPB) which
CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
CC purified from human plasma or by transformed cell culture by
CC extraction using plasminogen bound to a solid phase.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 423 AA;
SQ

Query Match 94.2%; Score 1799.5; DB 18; Length 423;
Best Local Similarity 89.6%; Pred. No. 3.6e-176;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
XX
QY 1 MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQVQLNLTYYEIVLMQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQVQLNLTYYEIVLMQPVTD 60
QY 61 LIVKKQVHFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASYY 120
DB 61 LIVKKQVHFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASYY 120
QY 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPLYLVKVGSKQTAKNAIWDGCI 180
DB 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPLYLVKVGSKQTAKNAIWDGCI 180
QY 181 HAREWISPAFLWFIH-----NRMRK 203
DB 181 HAREWISPAFLWFIH-----NRMRK 203
QY 204 NRSFYANNHCIGTDLNSNFVSKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 263
DB 241 NRSFYANNHCIGTDLNRFNFKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 300
QY 264 NOIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHSE 323
DB 301 NOIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHSE 360
QY 324 TLVLAPGGDDWIYDLGIKYSFT 346
DB 361 TLVLAPGGDDWIYDLGIKYSFT 383

RESULT 6
AAW92270
ID AAW92270 standard; Protein; 423 AA.
XX
XX AAW92270;
AC
XX 28-APR-1999 (first entry)
DT
XX
XX Human plasma carboxypeptidase B (PCPB) thr147.
DE
XX
XX Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease.
KW
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT Peptide 1...22
FT /note= "signal peptide"
FT Protein 23...423
FT /note= "mature PCPBthr147"
XX
XX W09855645-A1.
PN
XX
PD 10-DEC-1998.

XX 02-JUN-1998; 98WO-EP03244.
XX
XX 03-JUN-1997; 97US-0869057.
XX
XX (SCHD) SCHERING AG.
XX
XX Morser MJ, Nagashima M;
PI
XX WPI; 1999-045800/04.
XX
XX N-PSDB; AAV74302.
XX
XX
XX Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
XX Disclosure; Page 25-26; 35pp; English.
PS
XX
XX This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphisms of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphisms in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphisms within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
XX Sequence 423 AA;
SQ

Query Match 94.2%; Score 1799.5; DB 20; Length 423;
Best Local Similarity 89.6%; Pred. No. 3.6e-176;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
QY 1 MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQVQLNLTYYEIVLMQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQVQLNLTYYEIVLMQPVTD 60
QY 61 LIVKKQVHFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASYY 120
DB 61 LIVKKQVHFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASYY 120
QY 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPLYLVKVGSKQTAKNAIWDGCI 180
DB 121 EQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPLYLVKVGSKQTAKNAIWDGCI 180
QY 181 HAREWISPAFLWFIH-----NRMRK 203
DB 181 HAREWISPAFLWFIH-----NRMRK 203
QY 204 NRSFYANNHCIGTDLNSNFVSKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 263
DB 241 NRSFYANNHCIGTDLNRFNFKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 300
QY 264 NOIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHSE 323
DB 301 NOIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHSE 360
QY 324 TLVLAPGGDDWIYDLGIKYSFT 346
DB 361 TLVLAPGGDDWIYDLGIKYSFT 383

RESULT 7
AAR36273
ID AAR36273 standard; Protein; 423 AA.
XX
XX AAR36273;
AC

XX 24-AUG-1993 (first entry)
 XX Human plasma carboxypeptidase B.
 DE PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= signal_peptide
 FT Cleavage-site 114
 FT /note= "putative site of cleavage by trypsin to
 FT activate PCPB as a carboxypeptidase"
 FT Active-site 181
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 183
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 184
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 256..257
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 310
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 312
 FT /note= "expected to be involved in substrate
 FT binding"
 FT Active-site 369
 FT /note= "expected to be involved in catalytic
 FT activity"
 FT Binding-site 370..371
 FT /note= "expected to be involved in substrate
 FT binding; Asp 370 is the residue which
 FT determines specificity of the PCPB as
 FT a carboxypeptidase B"
 FT Binding-site 394
 FT /note= "expected to be involved in substrate
 FT binding"
 XX US206161-A.
 XX
 XX 27-APR-1993.
 XX
 XX 01-FEB-1991; 91US-0649591.
 XX
 XX 01-FEB-1991; 91US-0649591.
 XX (GETH) GENENTECH INC.
 XX Drayna DT, Eaton DL;
 XX WPI; 1993-151724/18.
 XX N-PSDB; AQA41001.
 XX
 XX New human plasma carboxypeptidase B - used as haemostatic
 XX regulator for clotting blood, partic. for treating blood clotting
 XX disorders, e.g. haemophilia
 XX Claim 2; Fig 4; 40pp; English.
 XX
 XX This amino acid sequence was deduced from a clone isolated from a
 XX human liver cDNA library. There is about 40% sequence identity
 XX between the deduced preproPCPB amino acid sequence and pro-human
 XX mast cell carboxypeptidase A and between preproPCPB and prepro-rat
 XX carboxypeptidase B. Human PCPB has the same substrate binding sites
 XX as, and shares 6 cysteine residues which form 3 intramolecular
 XX disulphide bonds with, bovine and rat carboxypeptidase B. The

CC presence of the same amino acid (Asp at position 348 of the mature
 CC protein) at the region in carboxypeptidases that determines
 CC substrate specificity, suggests that PCPB represents a plasma-
 CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
 CC tPA of plasminogen to plasmin in the presence of fibrinogen.
 XX
 XX SQ Sequence 423 AA;
 Query Match 93.9%; Score 1794.5; DB 14; Length 423;
 Best Local Similarity 89.3%; Pred. No. 1.2e-175; Indels 37; Gaps 1;
 Matches 342; Conservative 1; Mismatches 3;
 Qy 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQLNTTYYEIVLWQPVTD 60
 Db 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQLNTTYYEIVLWQPVTD 60
 Qy 61 LIVKKQVHFFVNASVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASY 120
 Db 61 LIVKKQVHFFVNASVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASY 120
 Qy 121 EYHSLNEIYSWIEFITERHPDMLTKIHGSSSEKYPYLVKVGKEQTAKNAIWDGCI 180
 Db 121 EYHSLNEIYSWIEFITERHPDMLTKIHGSSSEKYPYLVKVGKEQTAKNAIWDGCI 180
 Qy 181 HAREWISPAFLWFIGH-----NRMRK 203
 Db 181 HAREWISPAFLWFIGHITQFYGIQYTNLLRLVDYVNVVVDGYDYSWKKRMRK 240
 Qy 204 NRSFYANNHCIGTDLNSNFVSKWCEGASSSCSETYCGLYPESEPEVKAVASFLRNI 263
 Db 241 NRSFYANNHCIGTDLNRFNFKWCEGACSSCSETYCGLYPESEPEVKAVASFLRNI 300
 Qy 264 NQKAYISMHSYSHQIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
 Db 301 NQKAYISMHSYSHQIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360
 Qy 324 TLYLAPGGDDWIYDLGIKYSFT 346
 Db 361 TLYLAPGGDDWIYDLGIKYSFT 383
 RESULT 8
 AAB11458
 ID AAB11458 standard; Protein; 338 AA.
 XX
 AC AAB11458;
 XX
 DT 01-MAR-2001 (first entry)
 XX
 DE Human brain carboxypeptidase B protein SEQ ID NO 3.
 XX
 KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
 KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
 KW cerebroprotective; anti-Alzheimers; neurotropic; neuroprotective;
 KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
 KW Down's syndrome; head trauma.
 XX
 OS Homo sapiens.
 XX
 XX WO2000066717-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 01-MAY-2000; 2000WO-JP02878.
 XX
 PF 30-APR-1999; 99JP-0125169.
 XX
 PA (WATS/) MATSUMOTO A.
 XX
 PI Matsumoto A;
 XX
 DR WPI; 2000-687534/67.
 XX

PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
XX
PS Disclosure; Page 71-73; 84pp; Japanese.
XX
XX This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotropic, neuroprotective and hemostatic activity, and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 338 AA;
Query Match 93.4%; Score 1785; DB 21; Length 338;
Best Local Similarity 99.1%; Pred. No. 7.8e-175;
Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 23 FQSGVLAALPRTSRQVQLQNLTYYEIVLWQVTDADLVKKQVHFFVNASDNDVKA 82
DB 1 FQSGVLAALPRTSRQVQLQNLTYYEIVLWQVTDADLVKKQVHFFVNASDNDVKA 60
QY 83 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASYYEQVHSLNEIYSWIEFITERHPD 142
DB 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASYYEQVHSLNEIYSWIEFITERHPD 120
QY 143 MLTKIHIGSSFEKYPYLVKLVKSGKEQTAKNAIWDGCIHAREWISPAFCLEWFGHNRMR 202
DB 121 MLTKIHIGSSFEKYPYLVKLVKSGKEQTAKNAIWDGCIHAREWISPAFCLEWFGHNRMR 180
QY 203 KNRSFYANNHCIGTDLNSNFVSKHWCBEAGSSSCSTYCGLYPESPEVKAVASFLRRN 262
DB 181 KNRSFYANNHCIGTDLNRFASKHWCBEAGSSSCSTYCGLYPESPEVKAVASFLRRN 240
QY 263 INQIKAYISMHSYSQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 322
DB 241 INQIKAYISMHSYSQHIYVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 300
QY 323 ETLYLAPGGDDWIYDLGIKYFTSNPPVEKLLPLSLK 360
DB 301 ETLYLAPGGDDWIYDLGIKYFTSNPPVEKLLPLSLK 338
RESULT 9
AAB11459
ID AAB11459 standard; Protein; 246 AA.
XX
XX AAB11459;
XX
XX
DT 01-MAR-2001 (first entry)
XX
XX Human brain carboxypeptidase B protein SEQ ID NO 4.
XX
XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; neurotropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
XX Homo sapiens.
XX
XX WO200066717-A1.
XX
XX 09-NOV-2000.
XX
XX 01-MAY-2000; 2000WO-JP02878.
XX
XX 30-APR-1999; 99JP-0125169.
XX
XX

PA (MATS/) MATSUMOTO A.
XX
XX Matsumoto A;
XX
XX WPI; 2000-687534/67.
XX
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
XX Disclosure; Page 74-75; 84pp; Japanese.
XX
XX This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotropic, neuroprotective and hemostatic activity, and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 246 AA;
Query Match 69.3%; Score 1325; DB 21; Length 246;
Best Local Similarity 98.8%; Pred. No. 1.1e-127;
Matches 243; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 115 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKLVKSGKEQTAKNAI 174
DB 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKLVKSGKEQTAKNAI 60
QY 175 WIDCGIHAREWISPAFCLEWFGHNRMRKNSFYANNHCIGTDLNSNFVSKHWCBEAGSS 234
DB 61 WIDCGIHAREWISPAFCLEWFGHNRMRKNSFYANNHCIGTDLNRFASKHWCBEAGSS 120
QY 235 SSCSETYCGLYPESPEVKAVASFLRRNINQIKAYISMHSYSQHIYVFPYSYTRSKSKDHE 294
DB 121 SSCSETYCGLYPESPEVKAVASFLRRNINQIKAYISMHSYSQHIYVFPYSYTRSKSKDHE 180
QY 295 ELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYFTSNPPVEKL 354
DB 181 ELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYFTSNPPVEKL 240
QY 355 LPLSLK 360
DB 241 LPLSLK 246
RESULT 10
AAB58129
ID AAB58129 standard; Protein; 211 AA.
XX
XX AAB58129;
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polypeptide sequence SEQ ID 467.
XX
XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerable;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
XX Homo sapiens.
XX
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX

```
PR 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX Ruben SM;
XX WPI; 2000-587514/55.
DR N-PSDB; AAF18005.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX Claim 11; Page 944-945; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterization of the polynucleotide and protein
CC sequences.
XX
XX SQ Sequence 211 AA;
Query Match 42.0%; Score 803.5; DB 21; Length 211;
Best Local Similarity 88.8%; Pred. No. 4.6e-74;
Matches 151; Conservative 2; Mismatches 8; Indels 9; Gaps 2;
Qy 185 WISPAFCLWFI-----GHNRMWRKNRSPYANNHCIGTDLNFSKHWCEGASSSS 236
Db 3 WIS-MLCRWLMMVMYNSWKKNRMWRKNRSPYANNHCIGTDLNRFASKHWCEGASSS 61
Qy 237 CSETYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSQHIIVFPYSYTRSKSDHEEL 296
Db 62 CSETYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSQHIIVFPYSYTRSKSDHEEL 121
Qy 297 SLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346
Db 122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 171
RESULT 11
AAB99688
ID AAB99688 standard; Protein; 417 AA.
XX
AC AAB99688;
XX
XX 28-MAR-2003 (first entry)
XX
DE Amino acid sequence of human zinc carboxypeptidase B precursor.
XX Human; zinc carboxypeptidase B-like protein; cancer; CNS disorder;
KW chronic obstructive pulmonary disease; COPD; haematological disorder;
KW zinc carboxypeptidase B.
XX
XX Homo sapiens.
XX
XX WO200295018-A1.
XX
XX 28-NOV-2002.
XX
```

```
XX 16-MAY-2002; 2002WO-EP05392.
XX
XX 22-MAY-2001; 2001US-292329P.
XX
XX (FARB ) BAYER AG.
XX
XX Smolyar A;
XX
XX WPI; 2003-120801/11.
XX
XX New polynucleotide encoding a zinc carboxypeptidase B-like protein
PT antigens, useful for treating diseases associated with
PT carboxypeptidase dysfunction, e.g. cancer, CNS disorder, COPD or
PT haematological disorder -
XX
XX Disclosure; Fig 3; 113pp; English.
XX
XX The present sequence represents a zinc carboxypeptidase B precursor.
XX The specification describes a human zinc carboxypeptidase B-like
XX protein. The zinc carboxypeptidase B-like protein is used for treating
XX diseases, such as cancer, a CNS disorder, chronic obstructive pulmonary
XX disease (COPD) or haematological disorders. The zinc carboxypeptidase
XX B-like protein may also be used to identify compounds which may act
XX as activators or inhibitors at the enzyme's active site, to raise
XX specific antibodies which can block the enzyme and effectively reduce
XX its activity, as a bait protein in a two-hybrid or three-hybrid assay
XX to identify other proteins which bind to or interact with the human
XX zinc carboxypeptidase-like protein polypeptide and modulate its
XX activity, and for immunization of mammals.
XX
XX SQ Sequence 417 AA;
Query Match 33.5%; Score 641; DB 24; Length 417;
Best Local Similarity 35.3%; Pred. No. 7.7e-57;
Matches 135; Conservative 71; Mismatches 132; Indels 44; Gaps 7;
Qy 3 LCSLAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQLNLTITTYEIVLQPVTDL 61
Db 1 MLALLVLTVALASAHGGHFEGRVFRVNVDEHNHIIIRLASTQIDFWKPDSTQ 60
Qy 62 IVKKKQVHFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQIISNDTVSPRASASYE 121
Db 61 IKPHTVDFRVKAEDTTVTVENLVKQNELQYKVLISLNRNVVEAQFDSRV---RATGHSYE 117
Qy 122 QYHSLNIEYISWIFETERHDMLTKEHIGSSPKYPLVYLKVSQKEQTAKNAIWIDCGIH 181
Db 118 KYNKWETIEAWTQQVATENDPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFH 176
Qy 182 AREWISPAFCLWFI-----GH-----NRMWRKN 204
Db 177 AREWISPAFCWFVREAVRYGRIQVTELLNKLDFYVLVPLNIDGIYITWTWSRFRWKT 236
Qy 205 RSFYANNHCIGTDLNFSKHWCEGASSSSSETYCGLYPSEPEVKAVASFLRRNIN 264
Db 237 RSTHTGSSCIGTDPRNF-DAGWCEIGASRNPCEITYCGPAASEKETKALADFIRNKL 295
Qy 265 QIKAYISMHSYSQHIIVFPYSYTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
Db 296 SIKAYLTISYSQWMIYPYSYAYKLGNNALNALAKATVKEL-ASLHGKTYTYGPGATT 354
Qy 325 LYLAPGGDDWIYDLGIKYSFT 346
Db 355 IYPAAGSGDDWAYDQGIYSFT 376
RESULT 12
AAR75131
ID AAR75131 standard; Protein; 404 AA.
XX
XX AAR75131;
XX
XX 13-MAR-1996 (first entry)
XX
```

	Porcine Tyr-His-Met Procarboxypeptidase B.
KW	Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
KW	human serum albumin; premating factor alpha; mating factor alpha;
KW	proCBB.
XX	Sus scrofa.
OS	
XX	WO9514096-A1.
PN	26-MAY-1995.
XX	
PP	16-NOV-1994; 94WO-US31142.
XX	
PR	16-NOV-1993; 93US-0153258.
XX	(ELIL) LILLY & CO ELI.
PA	
PI	Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
PI	Sternner JL, Zhang H;
XX	
DR	WPI; 1995-200386/26.
DR	N-PSDB; AAQ90600.
XX	
PT	DNA encoding porcine carboxypeptidase B - used for transforming
PT	host cells, partic. Pichia species, for prodn. of the enzyme
XX	
PS	Example 1; Page 16-17; 34pp; English.
XX	
CC	The porcine carboxypeptidase B coding sequence can be placed in a
CC	bacterial or pref. Pichia yeast expression vector. The expression
CC	vector further comprises the signal peptide of either human serum
CC	albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC	(designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC	pFU474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC	(designated pLGD27 - NRRL B-21027). The method can be used for
CC	producing large amounts of porcine carboxypeptidase B and when
CC	produced in Pichia yeast, the protein does not need solubilization
CC	or folding. The produced enzyme is then used for pref. cleaving
CC	basic residues from the carboxy terminus of proteins.
XX	
SQ	Sequence 404 AA;
	Query Match 33.5%; Score 640.5; DB 16; Length 404;
	Best Local Similarity 37.7%; Pred. No. 8.3e-57;
	Matches 136; Conservative 60; Mismatches 122; Indels 43; Gaps 6
Qy	23 PQSGQVLALPRTSRQVQLNLTFTYEVLVLPQVPVTADLVKKKQVHFVNASDVNYKA 82
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy	10 PEGEKVFVRNVDENDISLLHELASTRQIDFWKPDSVTQIKPHSTVDPRVAEDILAVED 69
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy	83 HLNVSGTPCSVLADVDELIIQQOITSNDTVSPRASASYEQVHSLNEIVSWIEFITERHPD 142
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	70 FLEQNELQVEYLINLRSLVLAQFDSDRV---RTTGHSEYKYNNWTETIAEATWKQVTSNPD 126
Qy	143 MLTKKHIGSSPEKPYLVKVGREGQTAKNAIWDCGTHAREWISPAFLCFWI----- 195
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	127 LISTAGTTPLGNYYLLKV-KGPFGPKPALFCDFGHAREWISHAFQCMFVRRAVITY 185
Qy	196 -----GH-----NRMRKNRSFYANNHCITGTDLSNFVSK 225
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	186 GYESHMTEFLNKLDPFYVLPLVNLIDGYIYTWKRMWRKTRSTNACTTCIGTDPNENF-DA 244
Qy	226 HWCEGASSSSCSCTCYGLYPESEPEVKAVASFLLRRNIHQIKAYISMHSYSQHIVFPYSY 285
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	245 GWCTTGASTDPCDETTCGSAASEKETKALADFIRNLLSIKAYLIITHSYSQMILYPYSY 304
Qy	286 TRSKSKDHHEELSVLASEAVRALDKTSKNTRYTHGHSETLYLAPCGGDDWIYDLGIKYSF 345
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db	305 DYKUPPENNAELNLLAKAAVKEL-ATLYGTKYTYGGATTIYPAAAGSGDDWAYDQGIKYSF 363
Qy	346 T 346

Db 364 T 364

RESULT 13
AAW06172
XX ID AAW06172 standard; Protein; 415 AA.
XX AC
XX AAW06172;
XX DT 12-FEB-1997 (first entry)
XX DE Human pancreatic carboxypeptidase B.
XX KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
XX KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
XX KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
XX KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
XX KW Fd; F(ab')₂; PeIB; leader; human carboxypeptidase B.
XX OS Homo sapiens.
XX OS
XX FH Location/Qualifiers
FH Peptide 1..13
FT FT /note= "partial enzyme secretion leader sequence;
FT FT (pre-sequence)"
FT Peptide 14..108
FT FT /label= pro-sequence
FT Protein 109..415
FT FT /label= mature_enzyme
FT FT Misc-difference 243
FT FT /note= "not observed in other published HPCB
FT FT sequences"
FT Active-site 361
FT FT /note= "Asp253 characteristic of B type specificity"
XX W09620011-A1.
XX XX
XX PD 04-JUL-1996.
XX XX
XX PF 21-DEC-1995; 95WO-GB02991.
XX XX
XX PR 16-AUG-1995; 95GB-0016810.
XX PR 23-DEC-1994; 94GB-0026192.
XX XX
XX PA (ZENEC) ZENECA LTD.
XX XX
XX PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
XX PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
XX PI Tarragona-Fiol A, Taylorson CJ;
XX XX
XX DR WPI; 1996-321650/32.
XX DR N-PSDB; AAT42494.
XX XX
XX FT Two component system for anti-tumour therapy - comprising targeting
XX FT moiety linked to mutated enzyme which can transform an
XX FT anti-neoplastic prodrug
XX XX
XX FS Reference Example 14; Page 131-132; 182pp; English.
XX XX
XX CC A two-component system for anti-tumour therapy comprises a targeting
XX CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX CC prodrug. The system is based on antibody directed enzyme prodrug therapy
XX CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
XX CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
XX CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX CC be used. The present sequence is that of the native prepro-HCPB.
XX XX
XX SQ Sequence 415 AA;

Query Match 33.4%; Score 639; DB 17; Length 415;
Best Local Similarity 35.6%; Pred. No. 1.2e-56;
Matches 135; Conservative 69; Mismatches 131; Indels 44; Gaps

This Page Blank (uspto)

QY 592 ----- 591
Db 601 TTCTATGGGTAATAGGGCAATATACCAATCTCTGAGGCTTGGAATTTCTATGTTATG 660
QY 592 -----AATCGAATGTGAGAAAG 609
Db 661 CCGGTGTTAATGTGAGCGGTATGACTACTCATGGAAGAAAGATCGAATGTGAGAAAG 720
QY 610 AACCGTTCTTTCTATGCGAACAATCATTTGTCATGCGAAGACAGACCTGATAGCAACTTTTGTG 669
Db 721 AACCGTTCTTTCTATGCGAACAATCATTTGTCATGCGAAGACAGACCTGATAGCAACTTTTGTG 780
QY 670 TCMAACACCTGGTGTGAGGAAGGTGCATCCAGTTCTCTGTCGCGAAGACCTACTGTGGA 729
Db 781 TCMAACACCTGGTGTGAGGAAGGTGCATCCAGTTCTCTGTCGCGAAGACCTACTGTGGA 840
QY 730 CTTTATCTCAGTCAGAACAGAGTGAAGCGAGTGGCTAGTTCTTTGAGAAGAAATATC 789
Db 841 CTTTATCTCAGTCAGAACAGAGTGAAGCGAGTGGCTAGTTCTTTGAGAAGAAATATC 900
QY 790 AACCAAGTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTGTTTCCATAT 849
Db 901 AACCAAGTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTGTTTCCATAT 960
QY 850 TCCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCA 909
Db 961 TCCTATACAGAAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCA 1020
QY 910 GTTCGTCTATTGACAAACTAGTAAATACCAAGTATACATAGCCATGCGCTCAGAA 969
Db 1021 GTTCGTCTATTGACAAACTAGTAAATACCAAGTATACATAGCCATGCGCTCAGAA 1080
QY 970 ACCTTATACCTAGTCTCCTGAGGTGGGAGCATTTGGAATCTATGATTTGGGCATCAAAAT 1029
Db 1081 ACCTTATACCTAGTCTCCTGAGGTGGGAGCATTTGGAATCTATGATTTGGGCATCAAAAT 1140
QY 1030 TCGTTTACA 1038
Db 1141 TCGTTTACA 1149

RESULT 3

US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haack, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-07-649-591B-2

Query Match 84.2%; Score 909; DB 1; Length 1749;
Best Local Similarity 89.9%; Pred. No. 4.8e-270;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;
QY 1 ATGAAGCTTTGACGCTTGACGTCCTTGTAACCAATGTTCTCTCTGTGTGAGCAGCATGTC 60
Db 41 ATGAAGCTTTGACGCTTGACGTCCTTGTAACCAATGTTCTCTCTGTGTGAGCAGCATGTC 100
QY 61 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
Db 101 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 160
QY 121 GTTCTACAGAACTTACTACAAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGAC 180
Db 161 GTTCTACAGAACTTACTACAAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGAC 220
QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGATCTGTGATGTCGACAAATGTG 240
Db 221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGATCTGTGATGTCGACAAATGTG 280
QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGGCAGACGTTGGAAGAT 300
Db 281 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGGCAGACGTTGGAAGAT 340
QY 301 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCAGCCCGGAGCTCCCGCATCGTACTAT 360
Db 341 CTTATTCAACAGCAGATTTTCCAAACGACACAGTCAGCCCGGAGCTCCCGCATCGTACTAT 400
QY 361 GAAACAGTATCCTCACTAAATGAAATCTATTTCTTGGATAGAAATTTATACTGAGAGCAT 420
Db 401 GAAACAGTATCCTCACTAAATGAAATCTATTTCTTGGATAGAAATTTATACTGAGAGCAT 460
QY 421 CCTGATATGCTTACAAAAATCCACATTTGGATTCCTCATTTTGAGAAGTACCCACTCTATGTT 480
Db 461 CCTGATATGCTTACAAAAATCCACATTTGGATTCCTCATTTTGAGAAGTACCCACTCTATGTT 520
QY 481 TTAAGGTTTCTGGAAGAAAGCAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
Db 521 TTAAGGTTTCTGGAAGAAAGCAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580
QY 541 CATGCCAGAGAAATGGATCTCTCTGCTTTCTGTTGTTTCATAGGCCAT----- 591
Db 581 CATGCCAGAGAAATGGATCTCTCTGCTTTCTGTTGTTTCATAGGCCATATAACTCAA 640
QY 592 ----- 591

641	TTCTATGGGATATAGGGCAATATACCAATCTCTCTGAGGCTTGTGGATTTCTATGTTATG	700
592	-----AATCGAATCTGGAGAAAG	609
701	CCGGTGGTTAATGTGGACGGTTATGACTCTCATGGAAAAAGAAATCGAATGTGGAGAAAG	760
610	AACCGTTCTTTCTATGGCAAACAATCATTTGTCATCGGAAACAGACCTGGAATAGCAACTTTGTC	669
761	AACCGTCTTTCTATGGCAAACAATCATTTGTCATCGGAAACAGACCTGGAATAGCAACTTTGCT	820
670	TCCAAACACTGGTGTGAGGGAAGGTGCATCAGATTCTCTCATGCTCGGAAAACCTACTGTGGAA	729
821	TCCAAACACTGGTGTGAGGGAAGGTGCATCAGATTCTCTCATGCTCGGAAAACCTACTGTGGAA	880
730	CTTTATCTCAGTCAGAAACAGAAAGTGAAGGCACTAGTTCTTTTGAGAGAAGAAATATC	789
881	CTTTATCTCAGTCAGAAACAGAAAGTGAAGGCACTAGTTCTTTTGAGAGAAGAAATATC	940
790	AACCGATTAAAGCATACATCAGCATGCATTCTATCTCCAGCATATAGTGTGTTTCCATAT	849
941	AACCGATTAAAGCATACATCAGCATGCATTCTATCTCCAGCATATAGTGTGTTTCCATAT	1000
850	TCCTATACGAAAGTAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA	909
1001	TCCTATACGAAAGTAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA	1060
910	GTTTCGTGCTATTACAAAACCTAGTTAAAAATACACAGGTATACACATGCGCCATGCTCAGAA	969
1061	GTTTCGTGCTATTAGAAAACCTAGTTAAAAATACACAGGTATACACATGCGCCATGCTCAGAA	1120
970	ACCTTATACCTAGTCTCTCGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAAATAT	1029
1121	ACCTTATACCTAGTCTCTCGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAAATAT	1180
1030	TCGTTTACA	1038
1181	TCGTTTACA	1189

RESULT 4

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Db 821 TCCAACTGGTGTGAGGAGTGCATCCAGTTCCTCATCTCGGAACTCTAGTGG 880
Qy 730 CTTTATCTGAGTCAGAACAGAGTGAAGCAGTGTCTAGTCTTCTTGAAGAAATATC 789
Db 881 CTTTATCTGAGTCAGAACAGAGTGAAGCAGTGTCTAGTCTTCTTGAAGAAATATC 940
Qy 790 AACAGATTAAGCATACATCAGCATGCAATTCATATCCAGCATATAGTGTTCATAT 849
Db 941 AACAGATTAAGCATACATCAGCATGCAATTCATATCCAGCATATAGTGTTCATAT 1000
Qy 850 TCCTATACAGAACTAAAGCAAGACATGAGGAACTGTCTCTAGTACGAGTGAAGCA 909
Db 1001 TCCTATACAGAACTAAAGCAAGACATGAGGAACTGTCTCTAGTACGAGTGAAGCA 1060
Qy 910 GTTCGTCTATGTAAGAACTAGTAAATAACAGGTATACATGAGGCAATGCTCAGAA 969
Db 1061 GTTCGTCTATGTAAGAACTAGTAAATAACAGGTATACATGAGGCAATGCTCAGAA 1120
Qy 970 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCATCAATAT 1029
Db 1121 ACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCATCAATAT 1180
Qy 1030 TCGTTTACA 1038
Db 1181 TCGTTTACA 1189

RESULT 5
US-08-430-787A-2
; Sequence 2, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasek, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-430-787A-2

Query Match 84.2%; Score 909; DB 1; Length 1749;
Best Local Similarity 89.9%; Pred. No. 4.8e-270; Indels 111; Gaps 1;
Matches 1033; Conservative 0; Mismatches 5;

Qy 1 ATGAAGCTTTGCGAGCTTGCAGTCCCTGTGATCCCATTTCTCTCTCTGAGCAGCATGTC 60
Db |||||
Qy 41 ATGAAGCTTTGCGAGCTTGCAGTCCCTGTGATCCCATTTCTCTCTCTGAGCAGCATGTC 100
Db |||||
Qy 61 TTCCGCTTCCAGAGTGCCCAAGTTCCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 120
Db |||||
Qy 101 TTCCGCTTCCAGAGTGCCCAAGTTCCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 160
Db |||||
Qy 121 GTTCTACAGAACTTACTACAAATATGAGTGTCTCTCTGCGAGCCGCTAAACGCTGAC 180
Db |||||
Qy 161 GTTCTACAGAACTTACTACAAATATGAGTGTCTCTCTGCGAGCCGCTAAACGCTGAC 220
Db |||||
Qy 181 CTTATTGTGAAGAAAAAACAAGTCCCATTTTGTAAATGCACTGTGATGTCGACAAATGTG 240
Db |||||
Qy 221 CTTATTGTGAAGAAAAAACAAGTCCCATTTTGTAAATGCACTGTGATGTCGACAAATGTG 280
Db |||||
Qy 241 AAAGCCCATTTAAATGTGAGCGGAAATCCATGCAAGTGTCTTGTGGCAGACGTTGGAAGAT 300
Db |||||
Qy 281 AAAGCCCATTTAAATGTGAGCGGAAATCCATGCAAGTGTCTTGTGGCAGACGTTGGAAGAT 340
Db |||||
Qy 301 CTTATTCAACAGCAGATTTCCAAAGCAGACAGTCAAGCCCGGAGCTCCGATCGTACTAT 360
Db |||||
Qy 341 CTTATTCAACAGCAGATTTCCAAAGCAGACAGTCAAGCCCGGAGCTCCGATCGTACTAT 400
Db |||||
Qy 361 GAACAGTATCACTCACTAAATGAAATCTATTTTGGATGAAATTTATACTGAGAGCAT 420
Db |||||
Qy 401 GAACAGTATCACTCACTAAATGAAATCTATTTTGGATGAAATTTATACTGAGAGCAT 460
Db |||||
Qy 421 CTTGATATGCTTACAAAAATCCCATTTGGATCTCTCATTTGAGAAAGTACCCACTCTATGTT 480
Db |||||
Qy 461 CTTGATATGCTTACAAAAATCCCATTTGGATCTCTCATTTGAGAAAGTACCCACTCTATGTT 520
Db |||||
Qy 481 TTAAGGTTTCTGGAAGAAAGCAAAACAGCCAAAATGCCATATGGATTGATGTTGGAATC 540
Db |||||
Qy 521 TTAAGGTTTCTGGAAGAAAGCAAAACAGCCAAAATGCCATATGGATTGATGTTGGAATC 580
Db |||||
Qy 541 CATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGGTTTCATAGGCCAT----- 591
Db |||||
Qy 581 CATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGGTTTCATAGGCCATATNACTCAA 640
Db |||||
Qy 592 ----- 591
Db |||||
Qy 641 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTCTATGTTATG 700
Db |||||
Qy 592 -----AATCGAATGTGAGAAAG 609
Db |||||
Qy 701 CCGGTGGTTAATGTGGACGTTTATGACTACTCATGGAAGAAAGAAATCGAATGTGAGAAAG 760
Db |||||
Qy 610 AACCGTTCTTTCTATCGAACAAATCATTCATCGAACACAGACCTGATAGCACTTTGTC 669
Db |||||
Qy 761 AACCGTTCTTTCTATCGAACAAATCATTCATCGAACACAGACCTGATAGCACTTTGCT 820
Db |||||
Qy 670 TCCAAACACTGGTGTGAGGAAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db |||||
Qy 821 TCCAAACACTGGTGTGAGGAAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880
Db |||||
Qy 730 CTTTATCTGAGTCAGAAACAGAAAGTGAAGGAGTGTCTAGTCTTCTTGAAGAAATATC 789
Db |||||
Qy 881 CTTTATCTGAGTCAGAAACAGAAAGTGAAGGAGTGTCTAGTCTTCTTGAAGAAATATC 940
Db |||||
Qy 790 AACAGATTAAGCATACATCAGCATGCAATTCATATCCAGCATATAGTGTTCATAT 849
Db |||||
Qy 941 AACAGATTAAGCATACATCAGCATGCAATTCATATCCAGCATATAGTGTTCATAT 1000
Db |||||

QY 850 TCCTATACAGCAAGTAAAGCAAGACCATGAGAACTGCTCTAGTAGCCAGTGAAGCA 909
DB 1001 TCCTATACAGCAAGTAAAGCAAGACCATGAGAACTGCTCTAGTAGCCAGTGAAGCA 1060
QY 910 GTTCGTCTATTGACAAACCTAGTAAATAACAGGTATACACATGCCATGGCTCAGAA 969
DB 1061 GTTCGTCTATTGACAAACCTAGTAAATAACAGGTATACACATGCCATGGCTCAGAA 1120
QY 970 ACCTTATACCTAGTCTCTGAGGTGGGACGATGATGATGATGATGATGATGATGATGAT 1029
DB 1121 ACCTTATACCTAGTCTCTGAGGTGGGACGATGATGATGATGATGATGATGATGATGAT 1180
QY 1030 TCGTTTACA 1038
DB 1181 TCGTTTACA 1189

RESULT 6
US-09-813-133A-3
; Sequence 3, Application US/0981133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813.133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

Query Match 18.9%; Score 204; DB 4; Length 55827;
Best Local Similarity 100.0%; Pred. No. 4.1e-52;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAAGTGCATCCAGTCTCTCATGCTCGGAAACCTACTGTGGACATTTATCTGTAGTCA 744
DB 48643 GAGGAAGTGCATCCAGTCTCTCATGCTCGGAAACCTACTGTGGACATTTATCTGTAGTCA 48702
QY 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTTCGAGAGAAATATCAACCAAGTTAAAGCA 804
DB 48703 GAACCAAGTGAAGGAGTGGCTAGTTCTTTCGAGAGAAATATCAACCAAGTTAAAGCA 48762
QY 805 TACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTATACAGCAAGT 864
DB 48763 TACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTATACAGCAAGT 48822
QY 865 AAAAGCAAGACCATGAGCAACTG 888
DB 48823 AAAAGCAAGACCATGAGCAACTG 48846

RESULT 7
US-08-782-760-5
; Sequence 5, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:
; APPLICANT: Hartman, Jacob
; APPLICANT: Fulga, Netta
; APPLICANT: Mendelovitch, Simona
; APPLICANT: Gorecki, Marian
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas

; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,760
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/378,233
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..927
; US-08-782-760-5

Query Match 11.5%; Score 124; DB 2; Length 927;
Best Local Similarity 57.9%; Pred. No. 2.6e-28;
Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

QY 596 GAATGGGAGAAAGAACCGTTCTTTCTATGCGAACCAATCATTCGATCGGAACAGACCTGA 655
DB 362 GAATGGGAGAAAGAACCGCTCTACTATGCTGGAAGTTCTGCTGGGTGAGACCCCA 421
QY 656 ATAGCAACTTTGTCTCAAAACACTGGTGTGAGAAAGTGATCCAGTTCTCTATGCTCGG 715
DB 422 ACAGGAATTTTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478
QY 716 AARCTTACTGTGACCTTTATCTGAGTCAAGAACCAAGTGAAGGAGTGGCTAGTTTCT 775
DB 479 AARCTTACTGTGACCTTTATCTGAGTCAAGAACCAAGTGAAGGAGTGGCTAGTTTCT 538
QY 776 TGAGAGAAATATCAACCAAGTAAAGCATATCATGATGATGATGATGATGATGATGATGAT 835
DB 539 TCGCAACCACTCTCCACATCAAGGCTTACCTGACCATCCATCATCATCATCATCATGATGA 598
QY 836 TAGTGTTCATATTCCTATATACAGAGTAAAGCAAGCAACCATGAGGAACTGTCTCTAG 895
DB 599 TGCTCTACCTTACTCTCTATGACTCAAACTGCTGAGAACTATGAGGAAATGGAATGCC 658
QY 896 TAGCCAGTGAAGCAGTTCTGCTTATTCACAAAATAGTAAATATACAGGATATACATG 955
DB 659 TGGTGAAGGTGCGCAAGAGGCTTGCC---ACTGTGATGGCAACCAAGTACACATG 715
QY 956 GCCATGGCTCAGAAACCTTATATAGTCTCTGAGGTGGGACGATTTGATGATGATGAT 1015
DB 716 GCCAGAGCTACAACTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
QY 1016 TGGGCAATCAAAATATTCGTTTACATCAAAACCACTGTAGAGAGAGCTTTTCCGCTGTCTC 1075
DB 776 AGGAATCAAAATATTCGTTTACCTTTGA-ACCTCGGAGTACAGGCTTTCTTTGGCTTTCTC 834

RESULT 8

PCT-US96-00995-5
; Sequence 5, Application PC/TUS9600995
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00995
; FILING DATE: 25-JAN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..927
PCT-US96-00995-5

Query Match 11.5%; Score 124; DB 5; Length 927;

Best Local Similarity 57.9%; Pred. No. 2.6e-28;
Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

Qy	596	GAATGTGGAGAAAGAACCGTTCTTCTATGCGAACAAATCATTTGCGATCGGAACAGACCTGA	655
Db	362	GAATGTGGAGAAAGAACCGTTCTTCTATGCGAACAAATCATTTGCGATCGGAACAGACCTGA	421
Qy	656	ATAGCAACTTTGTTCTTCAAAACACTGGTGTGGAGAGGTGATCCAGTTCTCTATGCTGG	715
Db	422	ACAGGAATTTTAATGC---TGGCTGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG	478
Qy	716	AAACTTACTGTGACTTTATCTTCGATGTCAGACAGAGTGAAGGAGTGGCTAGTTTCT	775
Db	479	AACTTACTGTGAGACCGCCAGAGTCTGAAAAGAGACAAAGGCCCTGGCAGATTTC	538
Qy	776	TGAGAGAAATATCAACAGAGATTAAGCATACATCAGCATGCTATCTATCTCCAGCATA	835
Db	539	TCGCAACACCTTCTCACCATCAAGGCTTACCTGACCTCCACTCATATCTCAGATGA	598
Qy	836	TAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATAGGAAACTGCTCTAG	895
Db	599	TGCTCTACCTTACTCTCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAATGCC	658
Qy	896	TAGCCAGTGAAGCAGTTGCTGCTATTCAGAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	955
Db	659	TGTTGAAAGGTGGCGGAAAGGAGCTTGC---ACTCTGTCATGGCACCACCAAGTACACATATG	715

Qy	956	GCCATGGCTCAGAAACCTTATACCTAGTCTCTCGAGGTGGGACGATTTGATCTATGATT	1015
Db	716	GCCAGGAGCTACAACAATCTATCTCTGCTGCTGGGATCTGACGACTGCTTATGATC	775
Qy	1016	TGGGCATCAATATTCGTTTACATCAACCCACTCTAGAGAGCTTTTCCCGCTGTCTC	1075
Db	776	AGGAATCAATATTCCTTTACCTTTGA-ACTCGGATACAGGCTTCTTTTGGCTTTCTC	834

RESULT 9

US-08-696-139-3
; Sequence 3, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hershberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gavlo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..921
US-08-696-139-3

Query Match 11.4%; Score 123.4; DB 1; Length 921;

Best Local Similarity 56.3%; Pred. No. 3.9e-28;
Matches 273; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

Qy	555	GATCTCTCTGCTTTCTGCTTGTGGTTTCATAGCCCAATATCGAATGTGGAGAAAGAACCG	614
Db	321	GCTCAATATTGATGGCTACATCTACACCTGGACCAAGACCGAATGTGGAGAAAGAACCG	380
Qy	615	TTCTTTCTATGCGAAACAATCATTTGCGAAGACAGACCTGGAATAGCACTTTTCTCCAA	674
Db	381	CTCTACCAATGCTGGAACTACCTGCTGCACTGGCACAGACCCCAACAGAAATTTTGTGCTGG	440

QY 675 AACTGTGTCGAGGAGTGCATCCAGTTCCTCATGCTCGAAGACCTACTGTGACTTTA 734
Db 441 G---TGTGCAACTGAGGCTCTACAGACCCCTGCGATGAGACTTACTGTGATCTGC 497
QY 735 TCCTGAGTCAGAACCCAGAGAGTGAAGGAGTGGCTAGTTTCTTGAGAAAGAAATCAACCA 794
Db 498 TGCAGAGTCTGAAAAGAGACCAAGGCCCTGGCTGATTTTATAGCGCAACCTCTCCTC 557
QY 795 GATTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTA 854
Db 558 CATCAAGCATACCTGAGCATGCCACTCATATCAGAGATGATCTCTACCTTATTCTTA 617
QY 855 TACACGAGTAAAGAGCAAGACCATGAGGAAGTCTCTCTAGTAGCCAGTGAAGCAGTTGC 914
Db 618 TGATTACAAGTCTCCCGAGAACAAAGTCTGAGTTGAATACCTGGCTAAGGCTGCCGTGAA 677
QY 915 TGCTATTGACAAACTAGTAAATAACACAGATATACATAGGCCATGGCTCAGAAACCTT 974
Db 678 AGAAGTTG---CTACACTGTATGGCACCACCAAGTACACATACGGCCAGGAGTCAACAAT 734
QY 975 ATACCTAGTCTCTGGAGGTGGGAGCAGATTGATCTATGATTTGGGCATCAAAATTCGTT 1034
Db 735 CTATCTCTGCTCTGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAATATTCCTT 794
QY 1035 TACAT 1039
Db 795 CACCT 799

RESULT 10

US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hersberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; ~ LENGTH: 1215 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1215
US-08-696-139-1

Query Match 11.4%; Score 123.4; DB 1; Length 1215;
Best Local Similarity 56.3%; Pred. No. 4.5e-28;
Matches 273; Conservative 0; Mismatches 206; Indels 6; Gaps 2;
QY 555 GATCTCTCTGCTTCTGCTTGTGGTTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
Db 615 GCTCAATATTGATGGCTTACATCTACACCTGGACCAAGACCAGATGTGGAGAAAGAACCG 674
QY 615 TTCTTTCTATCGCAACAATCATTTGSCATCGGAACAGACCTGAATAGCAACTTTTGTCTCAA 674
Db 675 CTCTACCAATGCTGGAACCTACCTGCATTTGGCACAGACCCCAACAGAAATTTTGTGCTGG 734
QY 675 ACACCTGCTGAGGAAGGTGCATCCAGTTCTCATGCTCGGAAACCTACTGTGACTTTA 734
Db 735 G---TGTGCAACAAGTGGAGCTCTACAGACCCCTGCGATGAGACTTACTGTGATCTGC 791
QY 735 TCCTGAGTCAGAACCAAGTGAAGCAGTGGCTAGTTTCTTGGAGAAATATCAACCA 794
Db 792 TGAGAGTCTGAAAAGAGAGACCAAGGCCCTGGCTGATTTTATAGCAACAACCTCTCCTC 851
QY 795 GATTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTA 854
Db 852 CATCAAGCATACCTGACGATCCACTCATATCTACAGATGATACTCTACCTTATTCTTA 911
QY 855 TACACGAGTAAAGCAAGACCATGAGGAAGTGTCTCTAGTAGCCAGTGAAGCAGTTGC 914
Db 912 TGATTACAAACTCTCCCGAGAACAAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAA 971
QY 915 TGCTATTGACAAACTAGTAAATAACACAGTATACATAGGCCATGGCTCAGAAACCTT 974
Db 972 AGAAGTTG---CTACACTGTATGGCACCAGTACATACATGCGCCAGGAGTCAACAAT 1028
QY 975 ATACCTAGTCTCTGAGGTTGGGAGCAGTATGATTTGGGCATCAAAATTCGTT 1034
Db 1029 CTATCTGCTGCTGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAATATTCCTT 1088
QY 1035 TACAT 1039
Db 1089 CACCT 1093

RESULT 11

US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011,769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
;
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1263 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: other nucleic acid
; SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-011-769A-38

Query Match      10.1%; Score 109; DB 4; Length 1263;
Best Local Similarity 54.4%; Pred. No. 1.3e-23;
Matches 264; Conservative 0; Mismatches 215; Indels 6; Gaps 2;

Qy 555 GATCTCTCCTGCTTTCTGCTTGTGTTTCATAGGCCATAATCGAATGTGGAGAAAGACCG 614
Db |||||
Qy 645 GCTCAATATTGATGGCTACATCTACCTGACCAAGAGCGGATTTTGGAGAAAGACTCG 704
Db |||||
Qy 615 TTCTTTCTATCGGAACAATCATCTGCAGCAAGACCTGATAGCAACTTTGCTCCAA 674
Db |||||
Qy 705 CTCACCCACTACTGGATCTAGCTGCATTTGGACAGACCCCAACAGAAATTTGATGC--- 761
Db |||||
Qy 675 ACATGTGTGAGGAAGTGCATCCAGTTCTCTATGCTCGGAAACCTTACTGTGACTTTA 734
Db |||||
Qy 762 TGGTTGGTGTGAAATTTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGACCTGC 821
Db |||||
Qy 735 TCCTGAGTCAGACCAAGAGTGAAGGAGGCTAGTGTCTTGTGAGAAATATCAACCA 794
Db |||||
Qy 822 CGCAGAGCTGTGAAAGGAAACCAAGGCCCTGGCTGATTTTCATCGCAACAACTCTTTC 881
Db |||||
Qy 795 GATTAAGACATACATCAGCATGCTATCATCTCCAGCATATAGTGTTCATATTCCTA 854
Db |||||
Qy 882 CATCAGGCATATCTGACATCCACTGTACTCCCAATGATGATCTACCTTACTCAT 941
Db |||||
Qy 855 TACACGAAGTAAAGCAAGACCATGAGGAACTGTCTTAGTAGCCAGTGAAGCAGTTGC 914
Db |||||
Qy 942 TGCCTTACAACTCGTGTGAGAACATGCTGAGTTGAATGCCCTGCTAAAGCTACTGTGAA 1001
Db |||||
Qy 915 TGCTATTGACAAACTAGTAAATAATACAGTATACATAGTGGCCATGGCTCAGAAACCTT 974
Db |||||
Qy 1002 AGAATCTTGC---TCACCTGACCGCACCAAGTACACATATGCGCCGGAGCTACAACA 1058
Db |||||
Qy 975 ATACCTAGCTCTCGAGTGGGAGGATGATCTATGATTTGGGCATCAAAATTCGTT 1034
Db |||||
Qy 1059 CTATCTGCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGAAATCAGATATTCCTT 1118
Db |||||
Qy 1035 TACAT 1039
Db |||||
Qy 1119 CACCT 1123
Db |||||

```

RESULT 12

```

US-08-860-882A-67
; Sequence 67, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN

```

```

; APPLICANT: EGSELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 999 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-860-882A-67

```

```

Query Match      9.9%; Score 107.4; DB 2; Length 999;
Best Local Similarity 54.2%; Pred. No. 3.5e-23;
Matches 263; Conservative 0; Mismatches 216; Indels 6; Gaps 2;

Qy 555 GATCTCTCCTGCTTTCTGCTTGTGTTTCATAGGCCATAATCGAATGTGGAGAAAGACCG 614
Db |||||
Qy 387 GCTCAATATTGATGGCTACATCTACCTGACCAAGAGCGGATTTTGGAGAAAGACTCG 446
Db |||||
Qy 615 TTCTTTCTATCGGAACAATCATTTGCATCGGAACAGACCTGAATAGCAACTTTGCTCCAA 674
Db |||||
Qy 447 CTCACCCACTACTGGATCTAGCTGCATTTGGACAGACCCCAACAGAAATTTGATGC--- 503
Db |||||
Qy 675 ACATGTGTGAGGAAGTGCATCCAGTTCTCTATGCTCGGAAACCTTACTGTGACTTTA 734
Db |||||
Qy 504 TGGTTGGTGTGAAATTTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGACCTGC 563
Db |||||
Qy 735 TCCTGAGTCAGACCAAGAGTGAAGGAGGCTAGTGTTCCTGAGAGAAATATCAACCA 794
Db |||||
Qy 564 CGCAGAGCTGTGAAAGGAGACCAAGGCCCTGGCTGATTTTCATCGCAACAACTCTTTC 623
Db |||||
Qy 795 GATTAAGACATACATCAGCATGCTATCTCCAGCATATAGTGTTCATATTCCTA 854
Db |||||
Qy 624 CATCAGGCATATCTGACATCCACTGTGATCTCCCAATGATGATCTACCTTACTCAT 683
Db |||||
Qy 855 TACACGAAGTAAAGCAAGACCATGAGGAACTGTCTTAGTAGCCAGTGAAGCAGTTGC 914
Db |||||
Qy 684 TGCTTACAAACTCGTGTGAGAACCAATGCTGAGTTGAATGCCCTGCTTAAAGCTACTGTGAA 743
Db |||||
Qy 915 TGCTATTGACAAACTAGTAAATAATACAGGATATACACATGGCCATGGCTCAGAAACCTT 974
Db |||||

```

Db 744 AGAATTGCC---TCACTGACGCGCACCAAGTACACATATGTCGCCGCGGAGCTACACAAAT 800
Qy 975 ATACTAGCTCTCGAGTGGGACGATGGATCTATGATTTGGGCATCAATATTCGTT 1034
Db 801 CTATCTGCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860
Qy 1035 TACAT 1039
Db 861 CACCT 865

RESULT 13

US-09-011-769A-50

; Sequence 50, Application US/09011769A

; Patent No. 6436891

; GENERAL INFORMATION:

; APPLICANT: SLATER, Anthony M.

; BLAKEY, David C.

; DAVIES, David H.

; HENNAM, John F.

; HENNEQUIN, Laurent F.A.

; MARSHAM, Peter R.

; DOWELL, Robert I.

; TITLE OF INVENTION: Chemical Compounds

; NUMBER OF SEQUENCES: 87

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pillsbury Madison & Sutro, LLP

; STREET: 1100 New York Ave., N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 1.44 Mb disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MS Word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/011,769A

; FILING DATE: 13-Feb-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB96/01975

; FILING DATE: 13-AUG-1996

; APPLICATION NUMBER: GB 9612295.7

; FILING DATE: 12-JUN-1996

; APPLICATION NUMBER: GB 9611019.2

; FILING DATE: 25-MAY-1996

; APPLICATION NUMBER: GB 9516810.0

; FILING DATE: 16-AUG-1995

; INFORMATION FOR SEQ ID NO: 50:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 999 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: other nucleic acid

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..987

; FEATURE:

; NAME/KEY: mat_peptide

; LOCATION: 67..987

; SEQUENCE DESCRIPTION: SEQ ID NO: 50:

US-09-011-769A-50

Query Match 9.9%; Score 107.4; DB 4; Length 999;

Best Local Similarity 54.2%; Pred. No. 3.5e-23;

Matches 263; Conservative 0; Mismatches 216; Indels 6; Gaps 2;

Qy 555 GATCTCTCTGCTTTCTGCTTGTGTTGATAGGCCATAATCGAATGTGGAGAAAGACCG 614

Db 387 GCTCAATATTGATGGCTACATCTACACCTGACCAAGCGCGATTTTGGAGAAAGACTCG 446

Qy 615 TTCTTTCTATGCGAACAAATCATTGTCATCGGAACAGACCTGAATAGCAACTTTTGTCTCCAA 674
Db 447 CTCACCCCATCTAGGATCTAGCTGCAATGGCAGACACCCCAACAGAAATTTTGTATGC--- 503
Qy 675 ACACTGGTGTGAGGAAAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTTACTGTGGACTTTA 734
Db 504 TGGTTGGTGTGAAATTTGGAGCCTCTCGAAACCCCTGTGATGAAACTTTACTGTGGACCTGC 563
Qy 735 TCCTGAGTCAGAACCAAGAGTGAAGGAGTGGCTAGTTCTTCTGAGAAGAAATATCAACCA 794
Db 564 CGCAGAGTCTGAAAGAGGAGACCAAGGCCCTGGTGTGATTTTCATCCGCAACAACTCTCTTC 623
Qy 795 GATTAAGCATACATCAGCATGCTATTCATATCTCCAGCATATAGTGTTCCTCATATTCCTA 854
Db 624 CATCAAGGCATATCTGACATCCACTGCTACTCCCAATGATGATCTACCTTACTCATTA 683
Qy 855 TACACGAAGTAAAGCAAGACCATGAGGAACCTGTCTCTAGTAGCCAGTGGAAGCAGTTGC 914
Db 684 TGCTTACAACTCGGTGAGAACATGCTGAGTTGAATGCCCTGCTAAAGCTACTGTGAA 743
Qy 915 TGCTATTGACAAACTAGTAAATAATACAGGTATACACATGGCCATGGCTCAGAAACCTT 974
Db 744 AGAACTTGC---TCACTGCACGCGCACCAAGTACACATATGGCCCGGAGCTACAAAT 800
Qy 975 ATACCTAGCTCCTGGAGTGGGACGATGATGATCTATGTTTGGGCATCAATATTCGTT 1034
Db 801 CTATCTGCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860
Qy 1035 TACAT 1039
Db 861 CACCT 865

RESULT 14

US-08-860-882A-64

; Sequence 64, Application US/08860882A

; Patent No. 5985281

; GENERAL INFORMATION:

; APPLICANT: TAYLORSON, CHRISTOPHER JOHN

; APPLICANT: EGGELTE, HENDRIKUS JOHANNES

; APPLICANT: TARRAGONA-FIOL, ANTONIO

; APPLICANT: RABIN, BRIAN ROBERT

; APPLICANT: BOYLE, FRANCIS THOMAS

; APPLICANT: HENNAM, JOHN FREDERICK

; APPLICANT: BLAKELY, DAVID CHARLES

; APPLICANT: MARSHAM, PETER ROBERT

; APPLICANT: HEATON, DAVID WILLIAM

; APPLICANT: DAVIES, DAVID HUW

; TITLE OF INVENTION: CHEMICAL COMPOUNDS

; NUMBER OF SEQUENCES: 77

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PILLSBURY, MADISON & SUTRO

; STREET: 1100 NEW YORK AVENUE, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; COMPUTER: IBM compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/860,882A

; FILING DATE: JUNE 23, 1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: DONALD J. BIRD

; REGISTRATION NUMBER: 25,323

; REFERENCE/DOCKET NUMBER: 9901/238653

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3027

TELEFAX: (202) 822-0944	Score 107.4; DB 2; Length 1053;
TELEX: 6174627 CUSH	Best Local Similarity 54.28; Pred. No. 3.6e-23; Indels 6; Gaps 2;
INFORMATION FOR SEQ ID NO: 64:	Matches 263; Conservative 0; Mismatches 216;
SEQUENCE CHARACTERISTICS:	
LENGTH: 1053 bases	
TYPE: nucleic acid	
STRANDEDNESS: single	
TOPOLOGY: linear	
US-08-860-882A-64	

Query Match	9.98;	Score 107.4; DB 2; Length 1053;
Best Local Similarity	54.28;	Pred. No. 3.6e-23; Indels 6; Gaps 2;
Matches 263; Conservative	0;	Mismatches 216;

QY	555	GATCTCTCCTCGCTTTTCGTCTTTGTGGTTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG	614
DB	387	GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG	446
QY	615	TTCTTTTCTATGGCAACAATCATTTGCATCGGAAACAGACCTGAAATAGCAACTTTTGTCTCAA	674
DB	447	CTCCACCCTACTGGATCTAGCTGCTGCAATTTGGCAGACACCCCAACAGAAATTTTGATGC---	503
QY	675	ACACTGGTGTGGAAGGTGCATCCACTTCTCATGCTCGGAACCTACTGTGGACTTTA	734
DB	504	TGGTTGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTTACTGTGGACCTGC	563
QY	735	TCCTGAGTCAGAACCAAGCTGAAGGCAGTGGCTAGTTCCTTGAGAAAGAAATATCAACCA	794
DB	564	CGCAGAGTCTGAAAGAGAGACCAAGCCCTGGCTGATTTTCATCCGCAACAACCTCTCTTC	623
QY	795	GATTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTA	854
DB	624	CATCAAGGCATATCTGCAATCCACTCGTACTCCCAATGATGATCTACCTTTACTCATA	683
QY	855	TACACGAGTAAAGCAAGCAACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCAGTTCG	914
DB	684	TGCTTACAAACTCGGTGAGAACAACTGTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA	743
QY	915	TGCTATTGACAAACTAGTAAAAAATACAGGTATACACATGGCCATGGCTCAGAAAACCTT	974
DB	744	AGAACTTGCC---TCAGTCGACGGCACAAGTACACATATGGCCCGGAGGTACACAAT	800
QY	975	ATACCTAGTCTCTGGAGGTGGGGACGATGGATCTATGATTTGGGCAATCAATATTCGTT	1034
DB	801	CTATCTCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT	860
QY	1035	TACAT 1039	
DB	861	CACCT 865	

RESULT 15
US-09-011-769A-46
; Sequence 46, Application US/09011769A
; Patent No. 6436691
GENERAL INFORMATION:
APPLICANT: SLATER, Anthony M.
BLAKEY, David C.
DAVIES, David H.
HENNAM, John F.
HENNEQUIN, Laurent P.A.
MARSHAM, Peter R.
DOWELL, Robert I.
TITLE OF INVENTION: Chemical Compounds
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, LLP
STREET: 1100 New York Ave., N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:

Search completed: January 1, 2004, 15:41:02
Job time : 78.6472 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 13:56:00 ; Search time 372.077 Seconds
(without alignments)
10046.789 Million cell updates/sec

Title: US-09-980-881a-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagcttgacgcttgc.....ttttgcgctgtctctaaaa 1080

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	909	84.2	1728	13	US-09-980-107-2396
2	854.6	79.1	1272	10	Sequence 2396, Ap
3	441	40.8	1400	9	Sequence 1, Appl
4	331.2	30.7	416	10	Sequence 24, Appl
5	282.4	26.1	397	10	Sequence 14595, A
6	131	12.1	1547	13	Sequence 8425, Ap
7	131	12.1	1547	13	Sequence 265133, A
8	131	12.1	1547	13	Sequence 265134, A
9	131	12.1	1547	13	Sequence 265135, A
10	131	12.1	1547	13	Sequence 265136, A
11	131	12.1	1547	14	Sequence 265133, A
12	131	12.1	1547	14	Sequence 265134, A
13	131	12.1	1547	14	Sequence 265135, A
14	129.6	12.0	431	10	Sequence 265136, A
15	107.6	10.0	1254	15	US-09-917-800A-468
					US-10-229-546-3

16	107.6	10.0	1622	12	US-10-429-802-21	Sequence 21, Appl
17	107.6	10.0	1622	15	US-10-229-546-1	Sequence 1, Appl
18	107.6	10.0	1633	13	US-10-341-434-187	Sequence 187, App
19	101	9.4	1870	10	US-09-910-059-112	Sequence 112, App
20	101	9.4	2154	10	US-09-910-059-124	Sequence 124, App
21	94.6	8.8	1332	10	US-09-954-456-1141	Sequence 1141, Ap
22	92.8	8.6	1311	14	US-10-200-344-9	Sequence 9, Appl
23	92.8	8.6	1993	12	US-10-274-639-33	Sequence 33, Appl
24	92.8	8.6	2128	14	US-10-200-344-13	Sequence 13, Appl
25	91.6	8.5	936	15	US-10-106-698-1866	Sequence 1866, Ap
26	88.6	8.2	517	15	US-10-198-846-12472	Sequence 12472, A
27	88.6	8.2	1050	14	US-10-200-344-11	Sequence 11, Appl
28	87.6	8.1	1200	15	US-10-200-910-7	Sequence 7, Appl
29	83.6	7.7	945	15	US-10-200-910-3	Sequence 3, Appl
30	83.6	7.7	945	15	US-10-200-910-11	Sequence 11, Appl
31	80.4	7.4	451	15	US-10-060-036-3265	Sequence 3265, Ap
32	80	7.4	1125	9	US-09-888-615-2	Sequence 2, Appl
33	80	7.4	1332	16	US-10-176-306-75	Sequence 75, Appl
34	80	7.4	1603	16	US-10-176-308-73	Sequence 73, Appl
35	80	7.4	1826	13	US-10-252-157-453	Sequence 453, App
36	79.2	7.3	878	15	US-10-198-846-4746	Sequence 4746, Ap
37	78.4	7.3	741	14	US-10-200-344-5	Sequence 5, Appl
38	78.4	7.3	1603	16	US-10-176-306-73	Sequence 73, Appl
39	76.8	7.1	948	9	US-09-888-615-1	Sequence 1, Appl
40	76.8	7.1	1187	15	US-10-198-846-13457	Sequence 13457, A
41	72.8	6.7	1341	9	US-09-925-297-69	Sequence 69, Appl
42	71.6	6.6	1306	9	US-09-923-779-144	Sequence 144, App
43	71.6	6.6	1306	9	US-09-962-436-557	Sequence 557, App
44	70.6	6.5	1254	9	US-09-925-297-58	Sequence 58, Appl
45	70.6	6.5	1260	13	US-10-345-680-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1

US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Query Match	84.2%	Score	909	DB	10	Length	1728
Best Local Similarity	89.9%	Pred. No.	7.2e+266				
Matches	1033	Conservative	0	Mismatches	5	Indels	111
		Gaps	1				
Qy	1	ATGAAGCTTTCAGAGCTTCAGTCCCTTGTATCCCATTTCTCTCTGTGAGCAGATGTC	60				
Db	20	ATGAAGCTTTCAGAGCTTCAGTCCCTTGTATCCCATTTCTCTCTGTGAGCAGATGTC	79				
Qy	61	TTGCGGTTTCAGAGTGCAGTCCAGTTCTAGCTGCTTCTCTAGAACCTCTAGGCAAGTTCAA	120				
Db	80	TTGCGGTTTCAGAGTGCAGTCCAGTTCAGTCTGCTTCTCTAGAACCTCTCTAGGCAAGTTCAA	139				

QY 592 -----ANTCGAATGTCGAGAAAG 609
Db 661 CCACTGGTTAATGTGGATGTTATGACTATCTATGGAAGAAAGATCGAATGTGGAGAAAG 720
QY 610 AACCGTCTCTTCTATGCGAAACAATCAITGTGCATCGGAACAGACCTGGAATAGCAACTTTGTC 669
Db 721 AACCGTCTCTTCTATGCGAACAATCGTTGCATCGGAACAGACCTGGAACAGCACTTTGCG 780
QY 670 TCGAAACACTGGTGTGAGGAAGGTGCAATCCAGTTTCTCTATGCTCGGAAACCTACTGTGGA 729
Db 781 TCGAAACACTGGTGTGAGGAAGGTGCAATCCAGTTTCTCTATGCTCGGAAACCTACTGTGGA 840
QY 730 CTTTATCTCGAGTCAGACACAGAGTGAAGGCACTGAGTCTTCTTCTGAGAGAAATATC 789
Db 841 CTTTATCTCGAGTCAGACACAGAGTGAAGGCACTGAGTCTTCTTCTGAGAGAAATATC 900
QY 790 AACAGATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTTCCTCATAT 849
Db 901 AACACATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTTCCTCATAT 960
QY 850 TCTATACAGAAAGTAAAGCAAGACATGAGGAACCTGCTCTAGTAGCCAGTGAAGCA 909
Db 961 TCTATATCTGAAGCAAAAGCAAGACATGAGGAATGCTCTAGTAGCCAGTGAAGCA 1020
QY 910 GTTCGTCTATTGACAAACTAGTAAATAACAGGTATACACATGCGCATGCTCAGAA 969
Db 1021 GTTCGTCTATTGACAAACTAGTAAATAACAGGTATACACATGCGCATGCTCAGAA 1080
QY 970 ACCTTATACCTAGTCTCTCGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1029
Db 1081 ACCTTATACCTAGTCTCTCGAGGTGGGACGATTGGATCTATGATTTGGGCATCAAAATAT 1140
QY 1030 TCGTTTACA 1038
Db 1141 TCGTTTACA 1149

RESULT 3

US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (25)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Query Match 40.8%; Score 441; DB 9; Length 1400;
Best Local Similarity 98.9%; Pred. No. 2.7e-123;
Matches 444; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 590 AATATCGAATGTGAGAAAGAACCGTTCTTTCTATGCGAAACAATCAITGTGCATCGGAACAG 649
Db 177 AGAATCGAATGTGAGAAAGAACCGTTCTTTCTATGCGAAACAATCAITGTGCATCGGAACAG 236
QY 650 ACCTGAATAGCACTTTGCTCTCGAAACACTGCTGGTGTGAGGAAGGTGCAATCCAGTTCTCTCAT 709

Db 237 ACCTGAATAGGAACCTTTGCTTCAAACACTGGTGTGAGGAAGGTGCAATCCAGTTCTCTCAT 296
QY 710 GCTCGGAAACCTACTGTGGACTTTATTCCTGAGTCAGAACAGAGTGAAGGCACTGTGCTA 769
Db 297 GCTCGGAAACCTACTGTGGACTTTATTCCTGAGTCAGAACAGAGTGAAGGCACTGTGCTA 356
QY 770 GTTCTCTGAGAAAGAAATATCAACAGATTAAGCATACATCAGCATGCAATTCATCTCCC 829
Db 357 GTTCTCTGAGAAAGAAATATCAACAGATTAAGCATACATCAGCATGCAATTCATCTCCC 416
QY 830 AGCATATAGTGTTCCTATATTCCTATACAGAAAGTAAAGCAAAAGACCATGAGGAACCTGT 889
Db 417 AGCATATAGTGTTCCTATATTCCTATACAGAAAGTAAAGCAAAAGACCATGAGGAACCTGT 476
QY 890 CTCTAGTAGCCAGTGAAGCAGTTTCGTCTATTGACAAACTAGTAAATAACAGGTATATA 949
Db 477 CTCTAGTAGCCAGTGAAGCAGTTTCGTCTATTGAGAAACTAGTAAATAACAGGTATATA 536
QY 950 CACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCT 1009
Db 537 CACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGACGATTTGGATCT 596
QY 1010 ATGATTTGGGCATCAAAATATTCGTTTACA 1038
Db 597 ATGATTTGGGCATCAAAATATTCGTTTACA 625

RESULT 4

US-09-960-352-14595
; Sequence 14595, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14595
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10
US-09-960-352-14595

Query Match 30.7%; Score 331.2; DB 10; Length 416;
Best Local Similarity 87.3%; Pred. No. 3.8e-90;
Matches 363; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 150 GATTGTCTCTGCGAGCGGTAAACAGCTGACCTATTGTGAAGAAAGAAACAGTCCATTT 209
Db 1 GATTGTCTCTGCGAGCGGTAAACAGCTGACCTATTGTGAAGAAAGAAACAGTCCATTT 60
QY 210 TTTTGTAAATGCACTGTGTCGACAAATGTGAAGCCCAATTTAAATGTGAGCGGAATTC 269
Db 61 TTTTGTAAATGCACTGTGTCGACAAATGTGAAGCCCAATTTAAATGTGAGCGGAATTC 120
QY 270 ATGCACTGTCTCTGCGAGCGGTAAACAGCTGACCTATTGTGAAGAAAGAAACAGTCC 329
Db 121 ATTCAGGTCCTGTTGGAATAATGTGAAGATCTTATCCGCGACAGACATTTCCAATGACAC 180
QY 330 AGTCAGCCCCCGAGCCCTCCGCATCGTACTATGAACAGATATCACTCACTAAATGAAATCTA 389
Db 181 CATCAGCCCCCGGACATCTCTCTCTACTATGAACAGATATCACTCACTAAATGAGATCTA 240
QY 390 TTCTTGGATAGTAATTTATTAACCTGAGAGGCATCTGATATGCTTACAAAATCCACATGG 449
Db 241 TTCTTGGATAGTAATTTATGACTGAGCGGTATCTGATATGCTTGAAGAAATCCACATGG 300

```
Qy 450 ATCCTCATTTGAGAGTACCACCTCTATGTTTAAAGTTTCTGAAAAGCAACAAACAGC 509
Db 301 ATCCTCATACGAGAGTACCACCTTATGTTTAAAGTTTCTAATAAAGCAACAAAGGC 360
Qy 510 CAAAATGCCATATGGAATGACTGTGAATCCATGCCAGAGAAATGGATCTCTCTG 565
Db 361 CAAAATGCCATGGAATGACTGTGAATCCACGCCAGAGAGTGGATCTCTCTG 416

RESULT 5
US-09-960-352-8425
; Sequence 8425, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8425
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 36-LIB34-005-Q1-E1-A12
US-09-960-352-8425

Query Match 26.1%; Score 282.4; DB 10; Length 397;
Best Local Similarity 83.9%; Pred. No. 2.8e-75;
Matches 319; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 1 ATGAAGCTTTGACGCTTGTGAGTCTTGTATCCCATTTCTTCTCTGAGCAGCATGTC 60
Db 18 ATGAAGCTTTATAGCTTGGAGTCTTGTGCCACCGTCTGTTCTGTGGGAGCATGCC 77
Qy 61 TTCCGGTTCAGAGTGCCAAAGTTCTAGCTGCTTCTCTAGAACCTTAGGCAAGTTCAA 120
Db 78 TTCCGCTTTCAGAGGGGCGAGGTTTATCTGCTCTCTAGAACCTCCAGGCAAGTTCAA 137
Qy 121 GTTCTACAGATCTTACTACAAATGAGATGTTCTCTGCGAGCGGTAAAGCTGAC 180
Db 138 ATTCTGAGAGTGTACTACAAATGAGATGTTCTCTGCGAGCGGTAAAGCTGAA 197
Qy 181 CTATTGTGAAGAAAAACAAAGTCCATTTTGTAAATGATCTGATGTCGAAATGTG 240
Db 198 TATATTGTGAAGGATACGAGTCCATTTTGTGAATGATCTGATGTAAGCAATGTG 257
Qy 241 AAAGCCATTAAATGTAGCGGAATTCATGAGTGTCTTGTGGCAGACGTGGAAGAT 300
Db 258 AAAGCCATTAAATGTAGCGGAATTCATGAGTGTCTTGTGGGAAATGTGGAAGAT 317
Qy 301 CTTATTCAAGCAGAGATTTCCAAAGCAGCAGTCCAGCCCGAGCTCGCATCTACTAT 360
Db 318 CTTATCCGGCAGCAGATTTCCAAATGACACCATCAGCCCGGCGCATCTCTCTCTACTAT 377
Qy 361 GAACAGTATCACTCACTAAA 380
Db 378 GAACAGTATCACTCACTAAA 397

RESULT 6
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
```

```
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265133
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265133

Query Match 12.1%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 685 GAGAGAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGACTTTATCTGAGTCA 744
Db 1417 GAGAGAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGACTTTATCTGAGTCA 1476
Qy 745 GAACCAAGTGAAGCAGTGGCTAGTTCTTTCAGAGAAATATCAACCAAGATTAAAGCA 804
Db 1477 GAACCAAGTGAAGCAGTGGCTAGTTCTTTCAGAGAAATATCAACCAAGATTAAAGCA 1536
Qy 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547

RESULT 7
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265134
```

```
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

Query Match      12.1%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 685 GAGGAGGTCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAGGTCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 1536

Qy 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547

RESULT 8
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265136
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

Query Match      12.1%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 685 GAGGAGGTCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAGGTCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 1536

Qy 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547

RESULT 9
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265136
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

Query Match      12.1%; Score 131; DB 13; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 685 GAGGAGGTCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAGGTCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAAGCA 1536

Qy 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547

RESULT 10
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 265133
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-265133

Query Match      12.1%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 1476

QY 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547
```

```
RESULT 11
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 265134
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-265133

Query Match      12.1%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 1476

QY 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547
```

```
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-265134

Query Match      12.1%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 1476

QY 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547

RESULT 12
US-10-027-632-265135
; Sequence 265135, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 265135
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-265135

Query Match      12.1%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 1476

QY 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547
```

```
US-10-027-632-265135

Query Match      12.1%; Score 131; DB 14; Length 1547;
Best Local Similarity 100.0%; Pred. No. 8.7e-29;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 744
Db 1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTATCCTGAGTCA 1476

QY 745 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 804
Db 1477 GAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAAGAAATATCAACCAAGATTAAGCA 1536

QY 805 TACATCAGCAT 815
Db 1537 TACATCAGCAT 1547
```

RESULT 13

US-10-027-632-265136
; Sequence 265136, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265136
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1547)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

Query Match 12.1%; Score 131; DB 14; Length 1547;

Best Local Similarity 100.0%; Pred. No. 8.7e-29;

Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 685 GAGGAGGTGCATCCAGTCTCTATGCTCGGAACCTACTGTGGACTTTATCTCTGAGTCA 744
DB 1417 GAGGAAGGTGCATCCAGTCTCTATGCTCGGAACCTACTGTGGACTTTATCTCTGAGTCA 1476
QY 745 GAACCAAGTGAAGGCGAGTGGCTAGTTCTTGAGAGAATATCAACCAAGATTAAAGCA 804
DB 1477 GAACCAAGTGAAGGCGAGTGGCTAGTTCTTGAGAGAATATCAACCAAGATTAAAGCA 1536
QY 805 TATCATCAGCAT 815
DB 1537 TATCATCAGCAT 1547

RESULT 14

US-09-917-800A-468/c
; Sequence 468, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02

; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 468
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA998857
; NAME/KEY: misc_feature
; LOCATION: (1)...(431)
; OTHER INFORMATION: n = a or c or g or t
US-09-917-800A-468

Query Match 12.0%; Score 129.6; DB 10; Length 431;

Best Local Similarity 87.6%; Pred. No. 1.1e-28;

Matches 141; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 877 CATGAGGAAGTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAACTAGTAAA 936
DB 431 CATGAGGAAGTGTCTCTAGTAGCCAGTTCGTGCTATTGACAAACTAGTAAA 372
QY 937 AATACCAAGTATACATGCGCATGCTCAGAACCTATACCTAGTCTCTCGAGGTGG 996
DB 371 AACACCAAGTATACATGCGCATGCTCAGAACCTATACCTAGTCTCTCGAGGTGG 312
QY 997 GAGCATTTGATCTATGATTTGGGCATCAAAATATTCGTTTAC 1037
DB 311 GAGCATTTGATCTATGATTTGGGCATCAAAATATTCGTTTAC 271

RESULT 15

US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MP101-156P1RM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-229-546-3

Query Match 10.0%; Score 107.6; DB 15; Length 1254;

Best Local Similarity 55.8%; Pred. No. 1e-21;

Matches 250; Conservative 0; Mismatches 194; Indels 6; Gaps 2;


```
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF022AC110P1.
FEATURES
    source              1..971
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="CS0DF022YE21"
                        /tissue_type="FETAL BRAIN"
                        /dev_stage="fetal"
                        /clone_lib="Homo sapiens FETAL BRAIN"
                        /notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT      282 a   201 c   211 g   272 t      5 others
ORIGIN
Query Match      61.1%; Score 660.2; DB 9; Length 971;
Best Local Similarity 86.7%; Pred. No. 2.5e-174;
Matches 783; Conservative 3; Mismatches 6; Indels 111; Gaps 1;
QY      1 ATGAAGCTTTGCAGCTTGAGCTGCTTGTATCCCAATGTTCTTCTGTGAGCAGCATGTC 60
DB      65 AWGAAGCTTTGCAGCTTGAGCTGCTTGTATCCCAATGTTCTTCTGTGAGCAGCATGTC 124
QY      61 TTGCGGTTCCAGAGTGCCCAAGTCTAGCTGCTTCTTCTAGAACCTCTAGGCAAGTCAA 120
DB      125 TTGCGGTTCCAGAGTGCCCAAGTCTAGCTGCTTCTTCTAGAACCTCTAGGCAAGTCAA 184
QY      121 GTTCTACAGAACTTACTACACATATGAGATTGTTCTCTGGCAGCGCGTAACAGCTGAC 180
DB      185 GTTCTACAGAACTTACTACACATATGAGATTGTTCTCTGGCAGCGCGTAACAGCTGAC 244
QY      181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTGCAAAATGTG 240
DB      245 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTGCAAAATGTG 304
QY      241 AAAGCCCATTTAATGTGAGCGGAATTCATGACGATGTTCTGCTGGCAGACGTTGGAAGAT 300
DB      305 AAAGCCCATTTAATGTGAGCGGAATTCATGACGATGTTCTGCTGGCAGACGTTGGAAGAT 364
QY      301 CTTATTCAACAGCAGATTTTCCACGACACAGTACGAGCCCGAGCTCCGCAATCGTACTAT 360
DB      365 CTTATTCAACAGCAGATTTTCCACGACACAGTACGAGCCCGAGCTCCGCAATCGTACTAT 424
QY      361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATTAACCTGAGAGCAT 420
DB      425 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATTAACCTGAGAGCAT 484
QY      421 CTTGATATGCTTACAAAATCCCATTTGGATCTCTATTGAGAGATACCCACTTATGTT 480
DB      485 CTTGATATGCTTACAAAATCCCATTTGGATCTCTATTGAGAGATACCCACTTATGTT 544
QY      481 TTAAGGTTTCTGGAAGAACAAACAGCCAAAATGCCATATGATGATGCTGGAATC 540
DB      545 TTAAGGTTTCTGGAAGAACAAACAGCCAAAATGCCATATGATGATGCTGGAATC 604
QY      541 CATGCCAGAGAATGGATCTCTCTGCTTTCTGCTTGGTGGTTTCATAGCCCAT----- 591
DB      605 CATGCCAGAGAATGGATCTCTCTGCTTTCTGCTTGGTGGTTTCATAGCCCATATAACTCAA 664
QY      592 ----- 591
DB      665 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTTGGATTTCTATGTTATG 724
QY      592 -----RATCGAATGTGAGAAAG 609
DB      725 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAAGAAATCGAATGTGGAGAAAG 784
QY      610 AACCGTTCTTCTATGCGAAACAATCATTTGTCATCGAACACAGACCTGTAATAGCAACTTTGTC 669
```

```
Db      785 AACCGTTCTTCTATGCGAAACAATCATTTGCGAACAGACCTGTAATAGGAACTTTGCT 844
QY      670 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACTACTGTGGA 729
DB      845 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACTACTGTGGA 904
QY      730 CTTTATCTCTGAGTCAGAACACAGAAAGTGAAGCGAGTGGCTAGTTTCTTTGAGAAGAAATATC 789
DB      905 CTTTATCTCTGAGTCAGAACACAGAAAGTGAAGCGAGTGGCTAGTTTCTTTGAGAAGAAATATC 964
QY      790 AAC 792
DB      965 AAC 967

RESULT 2
LOCUS      AK004045
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length
enriched library, clone:110032P04 product:carboxypeptidase B2
(plasma), full insert sequence.
ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary system
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadoya, K., Matsuoka, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Stabli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
```


QY 850 TCCTATACGAACTAAAGCAAGACCATGAGGAACTGTCTCTAGTAGCAGTGAAGCA 909
 |||||
 Db 999 TCCTATACGAACTAAAGCAAGGACACGAAGACTGTCTCTAGTGGCAGGGAAGCA 1058
 |||||
 QY 910 GTTCGTGCTATTGACAAAACCTAGTAAATAACCAAGGTATACACATGGCCCTCAGAA 969
 |||||
 Db 1059 GTTCGTGCAATTGAAAGTATTATATAAAACACCAAGGTACACACGCGAGTGGCTCAGAA 1118
 |||||
 QY 970 ACCTTATACCTAGTCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGGCATCAAAATAT 1029
 |||||
 Db 1119 AGTTTATATCTAGTCTCTGAGGTCTTCTGACGATTTGGATCTATGATTTGGGCATCAAAATAT 1178
 |||||
 QY 1030 TCGTTTACA 1038
 |||||
 Db 1179 TCGTTTACA 1187
 |||||

RESULT 3
 BG618629 788 bp mRNA linear EST 18-APR-2001
 LOCUS 602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
 DEFINITION mRNA sequence.
 ACCESSION BG618629
 VERSION BG618629.1 GI:13670000
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 788)
 NTH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM1630 row: 0 column: 15
 High quality sequence stop: 760.

FEATURES
 source
 1. 788
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4767974"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_76"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
 SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGGCGCATATGCCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGGCGGCACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 BASE COUNT 222 a 157 c 169 g 240 t
 ORIGIN

Query Match 54.4%; Score 587.2; DB 10; Length 788;
 Best Local Similarity 99.5%; Pred No. 7.8e-154;
 Matches 589; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGAAGCTTTGCAGCTTGAGTCTTTGTATCCCATTTCTTCTCTGTGAGCAGCATGTC 60
 |||||
 Db 66 ATGAAGCTTTGCAGCTTGAGTCTTTGTATCCCATTTCTTCTCTGTGAGCAGCATGTC 125
 |||||

QY 61 TTCCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTCTAGAACCTCTAGCAAGTTCAA 120
 |||||
 Db 126 TTCCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTCTAGAACCTCTAGCAAGTTCAA 185
 |||||
 QY 121 GTTCTACAGAACTTTACTACAAACATATGAGATTGTTCTCTGGCAGCGCGTAACAGCTGAC 180
 |||||
 Db 186 GTTCTACAGAACTTTACTACAAACATATGAGATTGTTCTCTGGCAGCGCGTAACAGCTGAC 245
 |||||
 QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAAATGTG 240
 |||||
 Db 246 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAAATGTG 305
 |||||
 QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCACATGAGTGTCTTGTGGCAGAGTGGGAAGAT 300
 |||||
 Db 306 AAAGCCCATTTAAATGTGAGCGGAATTCACATGAGTGTCTTGTGGCAGAGTGGGAAGAT 365
 |||||
 QY 301 CTTATTCAACAGCAGAGATTTCCAAACGACACAGTCAGCGCCCGAGCCTCCGCATCGTACTAT 360
 |||||
 Db 366 CTTATTCAACAGCAGAGATTTCCAAACGACACAGTCAGCGCCCGAGCCTCCGCATCGTACTAT 425
 |||||
 QY 361 GAACAGTATCCTCAGTAAATGAAATCTATTTCTGGATAGAAATTTATACTGAGAGGCAT 420
 |||||
 Db 426 GAACAGTATCCTCAGTAAATGAAATCTATTTCTGGATAGAAATTTATACTGAGAGGCAT 485
 |||||
 QY 421 CCTGATATGCTTACAAAATCCACATTTGGATCCTCATTTTGCAGAGTACCCACTCTATGTT 480
 |||||
 Db 486 CCTGATATGCTTACAAAATCCACATTTGGATCCTCATTTTGCAGAGTACCCACTCTATGTT 545
 |||||
 QY 481 TTTAAAGTTTCTGGAAGAAACAAACAGCCAAAATATGCTATGATTTGATGTTGGAATC 540
 |||||
 Db 546 TTTAAAGTTTCTGGAAGAAACAAACAGCCAAAATATGCTATGATTTGATGTTGGAATC 605
 |||||
 QY 541 CATGCCAGAGAATGGATCTCTCTGCTTCTGCTTCTGCTTCTGCTTCTATAGGCCATA 592
 |||||
 Db 606 CATGCCAGAGAATGGATCTCTCTGCTTCTGCTTCTGCTTCTATAGGCCATA 657
 |||||

RESULT 4
 HSM089979 standard; RNA; EST; 702 BP.
 ID HSM089979
 XX
 AC BX496754;
 XX
 SV BX496754.1
 XX
 DT 09-MAY-2003 (Rel. 75, Created)
 DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
 XX
 DE Homo sapiens mRNA; EST DKFP77911529_r1 (from clone DKFP77911529)
 XX
 KW EST; expressed sequence tag.
 XX
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 XX [1]
 RN 1-702
 RP Ansoorge W., Krieger S., Regiert T., Rittmüller C., Schwager B.,
 RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
 RT Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
 RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
 XX
 CC This is the 5' sequence of the clone insert
 CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 CC Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 CC sequenced by EMBL (European Molecular Biology Laboratories,
 CC Heidelberg/Germany) within the cDNA sequencing consortium of
 CC the German Genome Project.
 CC No sl sequence available.
 CC This clone (DKFP77911529) is available at the RZPD in Berlin.

CA). Note: this is a NIH MGC Library."

BASE COUNT 209 a 161 c 153 g 215 t

Query Match 52.0%; Score 561.8; DB 10; Length 738;
Best Local Similarity 99.6%; Pred. No. 1.1e-146;
Matches 563; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 28 GTACCCATTGTTCTTCTGTGAGCAGCATGTTTCGGTTCCAGAGTGGCCAAAGTTCTTA 87
Db 1 GTACCCATTGTTCTTCTGTGAGCAGCATGTTTCGGTTCCAGAGTGGCCAAAGTTCTTA 60

Qy 88 GCTGCTCTTCTGAACTCTAGCAGATTTCAAGTCTTACAGATCTTACTACACATAT 147
Db 61 GCTGCTCTTCTGAACTCTAGCAGATTTCAAGTCTTACTACAGATCTTACTACACATAT 120

Qy 148 GAGATGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGAAAAACAAAGTCCAT 207
Db 121 GAGATGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTGTGAAGAAAAACAAAGTCCAT 180

Qy 208 TTTTGTGTAATGATCTGATGTGCAAAATGTGAAGCCCAATTTAAATGTGAGCGGAATT 267
Db 181 TTTTGTGTAATGATCTGATGTGCAAAATGTGAAGCCCAATTTAAATGTGAGCGGAATT 240

Qy 268 CCATGCACTGCTTGTGCGAGAGCTGCAAGATCTTATTCAACAGCAGATTTCCAAAGC 327
Db 241 CCATGCACTGCTTGTGCGAGAGCTGCAAGATCTTATTCAACAGCAGATTTCCAAAGC 300

Qy 328 ACAGTCAGCCCCGAGCTCCGATCTGATCAACAGATATCACTCACTAAATGAAATC 387
Db 301 ACAGTCAGCCCCGAGCTCCGATCTGATCAACAGATATCACTCACTAAATGAAATC 360

Qy 388 TATTCCTGATGAGATTTAACTGAGAGGATCTGATGCTTACAAAATCCCAATT 447
Db 361 TATTCCTGATGAGATTTAACTGAGAGGATCTGATGCTTACAAAATCCCAATT 420

Qy 448 GGATCCTCATTTGAGAGATCCCACTCTAGTTTAAAGTTTCTGAAAAGAACAAACA 507
Db 421 GGATCCTCATTTGAGAGATCCCACTCTAGTTTAAAGTTTCTGAAAAGAACAAACA 480

Qy 508 GCCAAAATGCCATATGAGTTGAGTGGATCCATCCAGAGATCGATCTCTCTGCT 567
Db 481 GCCAAAATGCCATATGAGTTGAGTGGATCCATCCAGAGATCGATCTCTCTGCT 540

Qy 568 TTCTGCTTGTGGTTTCATAGGCCATA 592
Db 541 TTCTGCTTGTGGTTTCATAGGCCATA 565

RESULT 8
AV646979
LOCUS AV646979 GLC Homo sapiens cDNA clone GLCATD06 3', mRNA sequence.
DEFINITION
ACCESSION
VERSION AV646979.1 GI:9867993
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z. and Han, Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

PUBMED 11752456

COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source

1..735
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCATD06"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 200 a 158 c 172 g 205 t
ORIGIN

Query Match 51.4%; Score 555.2; DB 9; Length 735;
Best Local Similarity 98.4%; Pred. No. 7.7e-145;
Matches 571; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

Qy 14 GCCTTGAGCTCTTGATACCAATGTTCTCTCTGTGAGCAGCATGTTCTCGGTTCCAGA 73
Db 5 GCCTTGCGGTCTTGATACCAATGTTCTCTCTGTGAGCAGCATGTTCTCGGTTCCAGA 64

Qy 74 GTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTCAAGTTCTACAGATC 133
Db 65 GTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTCAAGTTCTACAGATC 124

Qy 134 TTACTACAACATATGAGATTTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTCTGAAGA 193
Db 125 TTACTACAACATATGAGATTTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATTCTGAAGA 184

Qy 194 AAAAAACAAGTCCATTTTTTTGTAATGCACTGATGTCGCAATGTGAAAGCCCATTTAA 253
Db 185 AAAAAACAAGTCCATTTTTTTGTAATGCACTGATGTCGCAATGTGAAAGCCCATTTAA 244

Qy 254 ATGTGAGCGGAATTCATGCACTGCTTCTGTGAGCAGCTGGAAGATCTTATTCACAGC 313
Db 245 ATGTGAGCGGAATTCATGCACTGCTTCTGTGAGCAGCTGGAAGATCTTATTCACAGC 304

Qy 314 AGATTTCCAAACGACACAGTCAGCCCGAGCTCCGATCGTACTATGAACAGTATCACT 373
Db 305 AGATTTCCAAACGACACAGTCAGCCCGAGCTCCGATCGTACTATGAACAGTATCACT 364

Qy 374 CACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCATCTCTGATATGCTTA 433
Db 365 CACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCATCTCTGATATGCTTA 424

Qy 434 CAAAAATCCACATGGATCTCTCATTTTGAGAAAGTACCACCTCTATGTTTAAA-GGTTTCT 492
Db 425 CAAAAATCCACATTTGGATCTCTCATTTTGAGAAAGTACCACCTCTATGTTTAAAAGGGTCT 484

Qy 493 GGAAGAAGAACAAACAGCAAAAATGCCATATGGATTCAGCTGTGGAATCCATGCCAGAGAA 552
Db 485 GGAAGAAGAACAAACAGCAAAAATGCCATATGGATTCAGCTGTGGAATCCATGCCAGAGAA 544

Qy 553 TGGATCTCTCTCTGCTTTCTGCTTTGTTGGTTTCATAGGCCATA 592
Db 545 TGGATCTCTCTCTGCTTTCTGCTTTGTTGGTTTCATAGGCCATA 584

RESULT 9
CB157619
LOCUS CB157619 541 bp mRNA linear
DEFINITION K-EST0216807 L17N670205n1 Homo sapiens cDNA clone
L17N670205n1-12-A11 5', mRNA sequence.
ACCESSION CB157619
VERSION CB157619.1 GI:28142751

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 541)
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsaung@mail.kribb.re.kr
Plate: 12 row: A column: 11
High quality sequence stop: 541.
FEATURES
source
1..541
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-12-All"
/sex="F"
/lab_host="Top10P"
/clone_lib="L17N670205n1"
/note="Organ: Liver; Vector: pT7T3-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."
BASE COUNT 156 a 121 c 107 g 157 t
ORIGIN
Query Match 49.6%; Score 536.2; DB 14; Length 541;
Best Local Similarity 99.4%; Pred. No. 1.5e-139;
Matches 538; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 5 AGCTTTTCAGAGTTCCTTGTACCCATGTTCTCTGTGAGCAGCATGCTTCG 64
Db 1 AGCTTTTCAGAGTTCCTTGTACCCATGTTCTCTGTGAGCAGCATGCTTCG 60
Qy 65 CGTTCCAGAGTGGCCAAAGTTCAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAGTTC 124
Db 61 CGTTTCAGAGTGGCCAAAGTTCAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAGTTC 120
Qy 125 TACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAAACGCTGACCTTA 184
Db 121 TACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAAACGCTGACCTTA 180
Qy 185 TTGTGAAGAAACAAAGTTCATTTTGTGAATGATCTGATGTCGACAAATGTGAAG 244
Db 181 TTGTGAAGAAACAAAGTTCATTTTGTGAATGATCTGATGTCGACAAATGTGAAG 240
Qy 245 CCCATTTAAATGTAGCGGAATTCATGTCAGTGTCTTCTGGCAGAGCGTGAAGATCTTA 304
Db 241 CCCATTTAAATGTAGCGGAATTCATGTCAGTGTCTTCTGGCAGAGCGTGAAGATCTTA 300
Qy 305 TTCAACAGCAGATTTCCAAAGACACAGTCAGCCCGCCGAGCTCGGATCTGATGATGAC 364
Db 301 TTCAACAGCAGATTTCCAAAGACACAGTCAGCCCGCCGAGCTCGGATCTGATGATGAC 360
Qy 365 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTAATCTGAGAGGCATCCTG 424
Db 361 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTAATCTGAGAGGCATCCTG 420
Qy 425 ATATGCTTACAAAATCCACATTTGGATCCTCATTTTGAGAGTACCCACTATGTTTTAA 484
Db 421 ATATGCTTACAAAATCCACATTTGGATCCTCATTTTGAGAGTACCCACTATGTTTTAA 480

485 AGTTTCTGAAAAAGAACAAACAGCCAAATGCAATATGATGCTGCTGGAATCATG 544
481 AGTTTCTGAAAAAGAACAAACAGCCAAATGCAATATGATGCTGCTGGAATCATG 540
545 C 545
541 C 541
RESULT 10
LOCUS BG618813 750 bp mRNA linear EST 18-APR-2001
DEFINITION 602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767895 5',
mRNA sequence.
ACCESSION BG618813 GI:13670184
VERSION BG618813.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 750)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabps-f@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI630 row: k column: 22
High quality sequence stop: 622.
FEATURES
source
1..750
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4767895"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattggcc); Site 2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATATGACC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGGCAGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
Kb (range 1.0-4.0 Kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT 219 a 146 c 171 g 213 t 1 others
ORIGIN
Query Match 48.4%; Score 523.2; DB 10; Length 750;
Best Local Similarity 97.9%; Pred. No. 7.8e-136;
Matches 572; Conservative 0; Mismatches 8; Indels 4; Gaps 4;
Qy 1 ATGAGCTTTGACGCTTGCAGTCCCTGTACCCATGTTCTCTCTGTGAGCAGCATGTC 60
Db 11 ATGAGCTTTGACGCTTGCAGTCCCTGTACCCATGTTCTCTCTGTGAGCAGCATGTC 70
Qy 61 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTC 120
Db 71 TTGCGCTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTC 130
Qy 121 GTTCTACAGAAATCTTACTACAAATATGAGATTGTTCTCTGGCAGCCGGTAAACGCTGAC 180
Db 131 GTTCTACAGAAATCTTACTACAAATATGAGATTGTTCTCTGGCAGCCGGTAAACGCTGAC 190

QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCACTGATGTCGCAATGTG 240
 DB 191 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCACTGATGTCGCAATGTG 250
 QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCACATGAGTCTTTGCTGGCAGACGTGGAAGAT 300
 DB 251 AAAGCCCATTTAAATGTGAGCGGAATTCACATGAGTCTTTGCTGGCAGACGTGGAAGAT 310
 QY 301 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCCTCCGCAATCGTACTAT 360
 DB 311 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCCTCCGCAATCGTACTAT 370
 QY 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATACTGAGAGGCAT 420
 DB 371 GAACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATACTGAGAGGCAT 430
 QY 421 CCTGATATGCTT-ACAAAATCCACATGAG-ATCCTCATTTGAGAACTACCACTC-TAT 477
 DB 431 CCTGATATGCTT-ACAAAATCCACATTTGGCATCTCTCATTCGAGAACTAACCACTCGTAT 490
 QY 478 GTTTTAAAGGTTTCT-GGAAAAAGAAACAAACAGCCCAAAAATGCCATATGGATTGACTGTGG 536
 DB 491 GTTTTAAAGGTTCTCGGAAAGAAAGAACAGCAGGCAAAAATGCCATATGGATTGACTGTGG 550
 QY 537 AATCCATGCCAGAAATGAGATCTCTCTGCTTTCTGCTTTGCTGTGGT 580
 DB 551 AATCCATGCCAGAAATGAGATCTCTCTGCTTTCTGCTTTGCTGTGGT 594

RESULT 11
 BG568240
 LOCUS 602587149F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715797 5',
 DEFINITION mRNA sequence.
 ACCESSION BG568240
 VERSION BG568240.1 GI:13575893
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1. (bases 1 to 691)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Straubeberg, Ph.D.
 Email: csapsb@mail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLML at:
 http://image.llnl.gov
 Plate: LCM1563 row: a column: 14
 High quality sequence stop: 691.

FEATURES
 source
 1..691
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="IMAGE:4715797"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_76"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
 SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATATGCGC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCGGCGGCAGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT 204 a 143 c 144 g 200 t
 ORIGIN

Query Match 47.1%; Score 508.8; DB 10; Length 691;
 Best Local Similarity 98.5%; Pred. No. 8.5e-132;
 Matches 524; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 1 ATGAAGCTTTGCGAGCCTTGCGAGTCTTTGTACCCATTTGTTCTTCTTGTGAGCAGCATGTC 60
 DB 19 ATGAGCTTTGCGAGCCTTGCGAGTCTTTGTACCCATTTGTTCTTCTTGTGAGCAGCATGTC 78
 QY 61 TTCGCGTTCCAGAGTGGCCAAAGTTCTAGCTGCTTCTCTAGAACCTCTAGGCAAGTTCAA 120
 DB 79 TTCGCGTTCCAGAGTGGCCAAAGTTCTAGCTGCTTCTCTAGAACCTCTAGGCAAGTTCAA 138
 QY 121 GTTCTACAGAACTTCTACTCAACATATGAGATTTGTTCTCTGGCAGCGGTAACAGCTGAC 180
 DB 139 GTTCTACAGAACTTCTACTCAACATATGAGATTTGTTCTCTGGCAGCGGTAACAGCTGAC 198
 QY 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAAATGTG 240
 DB 199 CTTA-TGTGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAAATGTG 257
 QY 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGGCAGACGTGGAAGAT 300
 DB 258 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGGCAGACGTGGAAGAT 317
 QY 301 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCCTCCGCAATCGTACTAT 360
 DB 318 CTTATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCCTCCGCAATCGTACTAT 377
 QY 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATACTGAGAGGCAT 420
 DB 378 GAACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATACTGAGAGGCAT 437
 QY 421 CCTGATATGCTTACAAAATCCACATTTGGATCTCTCAATTTGAGAACTACCCACTTATGTT 480
 DB 438 CCTGATATGCTTACAAAATCCACATTTGGATCTCTCAATTTGAGAACTACCCACTTATGTT 497
 QY 481 TTTAAGGTTTCTGGAAGGAAACAGCCAAAATGCCATATGGATTGACT 532
 DB 498 TTTAAGGTTTCTGGAAGGAAACAGCCAAAATGCCATATGAACTCAAT 549

RESULT 12
 AV658390
 LOCUS 601 bp mRNA linear EST 16-JAN-2002
 DEFINITION GLC Homo sapiens cDNA clone GLCFMA03 3', mRNA sequence.
 ACCESSION AV658390
 VERSION AV658390.1 GI:9879404
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1. (bases 1 to 601)
 Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z. and Han, Z.
 Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 21625106
 MEDLINE
 PUBMED 11752456
 COMMENT
 Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919 (ex.45)

Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

```

FEATURES
source
Location/Qualifiers
1..601
/organism="Homo sapiens"
/mol_type="RNA"
/db_xref="taxon:9606"
/clone="GLCMA03"
/tissue_type="cornea"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBlueScript"

```

	XUOI			
BASE COUNT	167 a	136 c	118 g	177 t
OPIGIN				3 others

FEATURES
SOURCE

Query Match	46.4%	Score 501.2;	DB 9;	Length 601;
Best Local Similarity	99.0%;	Pred. No. 1.1e-129;		
Matches 503;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

```
source
1. 303
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKCAPH12"
/tissue_type="hepatocellula
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/note="Vector: pBluescript
XhoI"
```

BASE COUNT	161 a	130 c	116 g	176 t
ORIGIN				

Query Match	45.8%	Score	494.4	DB	9	Length	583
Best Local Similarity	98.8%	Pred. No.	8.8e-128				
Matches	498	Conservative	0	Mismatches	6	Indels	0
						Gaps	0

Qy	1	ATGAGCTTTGCGAGCTTGCAGTCCCTGTGTATCCCAATTGTTCTCTCTGTGACGACGATGTC	60
Db	16	ATGAAGCTTTTGCAGCCTTTGCAGTCCCTGTGTATCCCAATTGTTCTCTCTGTGACGACGATGTC	75
Qy	61	TTGCGCTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA	120
Db	76	TTGCGCTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA	135
Qy	121	GTTCTACAGAATCTTACTACAACTATGAGATTGTTCTTGCGCAGCCGGTAAACAGCTGAC	180
Db	136	GTTCTACAGAATCTTACTACAACTATGAGATTGTTCTTGCGCAGCCGGTAAACAGCTGAC	195
Qy	181	CTTATTCTGGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAATGTG	240
Db	196	CTTATTCTGGAAGAAAAAACAAGTCCATTTTGTGTAATGCATCTGATGTCGACAATGTG	255
Qy	241	AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGCGCAGAGCTGGAGAT	300
Db	256	AAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGCGCAGAGCTGGAGAT	315
Qy	301	CTTATTCAAAGCAGATTTTCCAAACGACACAGTCAAGCCCGCAGGCTTCGCACTGTACTAT	360
Db	316	CTTATTCAAAGCAGATTTTCCAAACGACACAGTCAAGCCCGCAGGCTTCGCACTGTACTAT	375
Qy	361	GAACAGTATCACTCACTAAATGAAATCTATTTCTTGGATAGAAATTTATACTGAGAGGCA	420
Db	376	GAACAGTATCACTCACTAAATGAAATCTATTTCTTGGATAGAAATTTATACTGAGAGGCA	435
Qy	421	CCTGATATGCTTACAAAATCCCATTTGGATGCTCATTTGAGAAGTACCCCACTCTATGTT	480
Db	436	CCTGATATGCTTACAAAATCCCATTTGGATGCTCATTTGAGAAGTACCCCACTCTATGTT	495
Qy	481	TTAAAGGTTTCTGGAAAAGAACAA	504
Db	496	TTAAAGGTTTCTGGAAACCAAGCA	519

RESULT 14

[illegible]

ACCESSION	AV693037
VERSION	AV693037.1
KEYWORDS	GI:10294900
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 583)
AUTHORS	Xu, X., Huang, J., Xu, Z., Qian, B., Cai, T., Zhang, X., Xiao, H., Qiu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,

AV694058
LOCUS AV694058 623 bp mRNA linear EST 16-JAN-2002
DEFINITION AV694058 GK Homo sapiens cDNA clone GKCADB09 5', mRNA sequence.
ACCESSION AV694058
VERSION AV694058.1 GI:10295921
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 623)
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z. and Han, Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
JOURNAL
MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
Location/Qualifiers
1..623
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKCADB09"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 166 a 137 c 130 g 186 t 4 others
ORIGIN
Query Match 45.8%; Score 494.4; DB 9; Length 623;
Best Local Similarity 98.8%; Pred. No. 9.1e-128;
Matches 498; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 ATGAAGCTTTGACGCTTGTGAGTCTTGTACCCATTTCTTCTCTGTGAGCAGCATGTC 60
Db 17 ATGAAGCTTTGACGCTTGTGAGTCTTGTACCCATTTCTTCTCTGTGAGCAGCATGTC 76
Qy 61 TTGCGGTTCCAGAGTGCCCAAGTCTAGCTCTTCTCTAGACCTCTAGCAAGTTCAA 120
Db 77 TTGCGGTTCCAGAGTGCCCAAGTCTAGCTCTTCTCTAGACCTCTAGCAAGTTCAA 136
Qy 121 GTTCTACAGATCTTACTACAAATATGAGATTCTTCTTGGATAGAAATTTATACTGAGGCAT 180
Db 137 GTTCTACAGATCTTACTACAAATATGAGATTCTTCTTGGATAGAAATTTATACTGAGGCAT 196
Qy 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAATGTG 240
Db 197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAATGTG 256
Qy 241 AAAGCCCATTTAAATGTGAGGGGAATTCATGAGTGTCTTCTGCGCAGACGTTGGAAGAT 300
Db 257 AAAGCCCATTTAAATGTGAGGGGAATTCATGAGTGTCTTCTGCGCAGACGTTGGAAGAT 316
Qy 301 CTTATTCAACAGCAGATTTTCAACGACACAGTCAAGCCCCGAGCTCCGCGATCGTACTAT 360
Db 317 CTTATTCAACAGCAGATTTTCAACGACACAGTCAAGCCCCGAGCTCCGCGATCGTACTAT 376
Qy 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGGCAT 420

Db 377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGGCAT 436
Qy 421 CCTGATATGCTTACAAAAATCCACATTTGGATCTCTATTTGAGAAAGTACCCACTCTATGTT 480
Db 437 CCTGATATGCTTACAAAAATCCACATTTGGATCTCTATTTGAGAAAGTACCCACTCTATGTT 496
Qy 481 TTAAGGTTTCTCGAAAGAACAA 504
Db 497 TTAAGGTTTCTCGGAACCAAGCCA 520
RESULT 15
LOCUS AV692032 602 bp mRNA linear EST 16-JAN-2002
DEFINITION AV692032 GK Homo sapiens cDNA clone GKCACD01 5', mRNA sequence.
ACCESSION AV692032
VERSION AV692032.1 GI:10293895
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z. and Han, Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
JOURNAL
MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
Location/Qualifiers
1..602
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKCACD01"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 165 a 133 c 120 g 182 t 2 others
ORIGIN
Query Match 45.7%; Score 493.4; DB 9; Length 602;
Best Local Similarity 98.8%; Pred. No. 1.7e-127;
Matches 497; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1 ATGAAGCTTTGACGCTTGTGAGTCTTGTACCCATTTCTTCTCTGTGAGCAGCATGTC 60
Db 17 ATGAAGCTTTGACGCTTGTGAGTCTTGTACCCATTTCTTCTCTGTGAGCAGCATGTC 76
Qy 61 TTGCGGTTCCAGAGTGCCCAAGTCTAGCTGTCTTCTAGAACCTCTAGCAAGTTCAA 120
Db 77 TTGCGGTTCCAGAGTGCCCAAGTCTAGCTGTCTTCTAGAACCTCTAGCAAGTTCAA 136
Qy 121 GTTCTACAGATCTTACTACAAATATGAGATTCTTCTTGGATAGAAATTTATACTGAGGCAT 180
Db 137 GTTCTACAGATCTTACTACAAATATGAGATTCTTCTTGGATAGAAATTTATACTGAGGCAT 196

```
Qy 181 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 240
Db |||||||
Qy 197 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTCGACAAATGTG 256
Db |||||||
Qy 241 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGCTGGCAGACGTGGAAGAT 300
Db |||||||
Qy 257 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGCTGGCAGACGTGGAAGAT 316
Db |||||||
Qy 301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCGCATCGTACTAT 360
Db |||||||
Qy 317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCGCATCGTACTAT 376
Db |||||||
Qy 361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATAACTGAGAGGCAT 420
Db |||||||
Qy 377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAATTTATAACTGAGAGGCAT 436
Db |||||||
Qy 421 CCTGATATGCTTACAAAAATCCACATTTGGATCCTCATTTGAGAAAGTACCCACTCTATGTT 480
Db |||||||
Qy 437 CCTGATATGCTTACAAAAATCCACATTTGGATCCTCATTTGAGAAAGTACCCACTCTATGNT 496
Db |||||||
Qy 481 TTAAGGTTTCTGGAAGAACRA 504
Db |||||||
Qy 497 TTAAGGTTTCTGGAAGAACRA 520
Db |||||||
```

Search completed: January 1, 2004, 15:37:50
Job time : 2504.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:21:46 ; Search time 22.1712 Seconds
(without alignments)
687.014 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLPVILFCEQHV.....IKYSFTSNPPVEXKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1817	95.1	386	4	US-09-813-133A-2
2	1799.5	94.2	423	1	US-07-649-591B-3
3	1799.5	94.2	423	1	US-08-277-540-8
4	1799.5	94.2	423	1	US-08-430-787A-3
5	1799.5	94.2	423	2	US-08-869-057-2
6	1788.5	93.6	423	4	US-09-813-133A-4
7	640.5	33.5	404	1	US-08-696-139-2
8	639	33.4	415	2	US-08-860-882A-57
9	639	33.4	415	4	US-09-011-769A-39
10	631.5	33.0	424	4	US-09-011-769A-56
11	628.5	32.9	417	1	US-07-649-591B-7
12	628.5	32.9	417	1	US-08-277-540-7
13	628.5	32.9	417	1	US-08-430-787A-7
14	623	32.6	716	3	US-09-171-945-125
15	619.5	32.4	417	1	US-07-649-591B-6
16	619.5	32.4	417	1	US-08-277-540-6
17	619.5	32.4	417	1	US-08-430-787A-6
18	604	31.6	307	2	US-08-782-760-6
19	604	31.6	307	5	PCT-US96-00995-6
20	604	31.6	437	4	US-09-675-305-10
21	603.5	31.6	396	1	US-07-649-591B-4
22	603.5	31.6	396	1	US-08-277-540-4
23	603.5	31.6	396	1	US-08-430-787A-4
24	576	30.1	306	1	US-08-696-139-4
25	561	29.4	329	4	US-09-011-769A-51
26	561	29.4	349	4	US-09-011-769A-47
27	554	29.0	349	4	US-09-011-769A-60

28	553	28.9	349	4	US-09-011-769A-64	Sequence 64, Appl
29	553	28.9	417	1	US-07-649-591B-8	Sequence 8, Appl
30	553	28.9	417	1	US-08-277-540-8	Sequence 8, Appl
31	553	28.9	417	1	US-08-430-787A-8	Sequence 8, Appl
32	546	28.6	399	4	US-09-710-099-8	Sequence 8, Appl
33	543	28.4	613	3	US-09-171-945-113	Sequence 113, App
34	528.5	27.7	350	4	US-09-675-305-12	Sequence 12, Appl
35	528.5	27.7	419	1	US-07-649-591B-5	Sequence 5, Appl
36	528.5	27.7	419	1	US-08-277-540-5	Sequence 5, Appl
37	528.5	27.7	419	1	US-08-430-787A-5	Sequence 5, Appl
38	519.5	27.2	436	4	US-09-710-099-6	Sequence 6, Appl
39	513	26.8	417	3	US-08-640-906-4	Sequence 4, Appl
40	513	26.8	417	4	US-09-395-936-4	Sequence 4, Appl
41	512	26.8	417	3	US-08-640-906-18	Sequence 18, Appl
42	512	26.8	417	4	US-09-395-936-18	Sequence 18, Appl
43	487.5	25.5	419	3	US-08-640-906-2	Sequence 2, Appl
44	487.5	25.5	419	4	US-09-395-936-2	Sequence 2, Appl
45	482.5	25.2	419	3	US-08-640-906-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CLO01173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Query Match		95.1%	Score 1817;	DB 4;	Length 386;
Best Local Similarity		98.6%	Pred. No. 1e-188;		
Matches 341;		Conservative	1;	Mismatches 4;	Indels 0; Gaps 0;
Qy	1	MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPV	60		
Db	1	MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPV	60		
Qy	61	LIVKKQVHFVNASDVNDVKAHNVSGIPCSVLLADVEDLIQQOISNDTVSPRAS	120		
Db	61	LIVKKQVHFVNASDVNDVKAHNVSGIPCSVLLADVEDLIQQOISNDTVSPRAS	120		
Qy	121	EQYHSLNEIYSWTEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKNAI	180		
Db	121	EQYHSLNEIYSWTEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKNAI	180		
Qy	181	HAREWISPAFLWFIQHNRMWRKNSFYANNHCIGTDLNSFNFKWCEGASSSS	240		
Db	181	HAREWISPAFLWFIQHNRMWRKNSFYANNHCIGTDLNSFNFKWCEGASSSS	240		
Qy	241	YCLGLYPESEBEVKAVASFLRRNINQIKAYISMHSYSHIVFPYSYTRSKSKD	300		
Db	241	YCLGLYPESEBEVKAVASFLRRNINQIKAYISMHSYSHIVFPYSYTRSKSKD	300		
Qy	301	SEAVRAIDKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSFT	346		
Db	301	SEAVRAIDKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSFT	346		

RESULT 2

US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: NO. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Query Match 94.2%; Score 1799.5; DB 1; Length 423;
Best Local Similarity 89.6%; Pred. No. 9.7e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
Qy 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQLTTTYYEIVLWQPVTD 60
Db 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQLTTTYYEIVLWQPVTD 60
Qy 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASASY 120
Db 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASASY 120
Qy 121 EQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSKQAKNAIWDGCI 180
Db 121 EQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSKQAKNAIWDGCI 180
Qy 181 HAREWISPAFLMFIH-----NRMRK 203
Db 181 HAREWISPAFLMFIH-----NRMRK 203
Qy 204 NRSFYANNHCIGTDLNSNFVSKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 263
Db 241 NRSFYANNHCIGTDLNRFASKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 300
Qy 264 NQKAYISMHSYSQHIYFPYSTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHS 323
Db 301 NQKAYISMHSYSQHIYFPYSTRSKSDHEELSLVASEAVRALEKTSKNTRYTHGHS 360
Qy 324 TLYLAPGGDDWIYDLGKYSFT 346
Db 361 TLYLAPGGDDWIYDLGKYSFT 383

RESULT 3
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689DIC1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-277-540-3

Query Match 94.2%; Score 1799.5; DB 1; Length 423;
Best Local Similarity 89.6%; Pred. No. 9.7e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
Qy 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQLTTTYYEIVLWQPVTD 60
Db 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQLTTTYYEIVLWQPVTD 60
Qy 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASASY 120
Db 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASASY 120
Qy 121 EQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSKQAKNAIWDGCI 180
Db 121 EQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSKQAKNAIWDGCI 180
Qy 181 HAREWISPAFLMFIH-----NRMRK 203
Db 181 HAREWISPAFLMFIH-----NRMRK 203
Qy 204 NRSFYANNHCIGTDLNSNFVSKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 263
Db 241 NRSFYANNHCIGTDLNRFASKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLRNI 300

QY 264 NOIKAYISMHSYSHOIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
DB 301 NOIKAYISMHSYSHOIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360
QY 324 TLYLAPGGDDWIYDLGKYSFT 346
DB 361 TLYLAPGGDDWIYDLGKYSFT 383

RESULT 4

US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA: US/08/430,787A
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994

; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91

; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.

; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1

; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881

; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

US-08-430-787A-3

Query Match 94.2%; Score 1799.5; DB 1; Length 423;
Best Local Similarity 89.6%; Pred. No. 9.7e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVILFCEQHVFAFOSGOVLAALPRTSRQVQLQNLTTTYEIVLWQPTAD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFOSGOVLAALPRTSRQVQLQNLTTTYEIVLWQPTAD 60

QY 61 LIVKKQVHFVFNASDVNDVNAHLNVSGIPCSVLLADVEDLIQQISNDTVSPRASASY 120
DB 61 LIVKKQVHFVFNASDVNDVNAHLNVSGIPCSVLLADVEDLIQQISNDTVSPRASASY 120

QY 121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKVGKEQTAKNAIWDGCI 180
DB 121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKVGKEQTAKNAIWDGCI 180

DB 121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPVLYLVKVGKEQTAKNAIWDGCI 180
QY 181 HAREWISPAFLWFIH-----NRMRK 203
DB 181 HAREWISPAFLWFIHITQFYGIQYTNLLRLVDFYVMPVNVVDGYDYSWKKNRMRK 240
QY 204 NRSFYANNHCIGTDLNSNFVSKWCEGASSSCSETYCGLYPESEBEVKAVASFLRNI 263
DB 241 NRSFYANNHCIGTDLNRNPFASKWCEGASSSCSETYCGLYPESEBEVKAVASFLRNI 300
QY 264 NOIKAYISMHSYSHOIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
DB 301 NOIKAYISMHSYSHOIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360
QY 324 TLYLAPGGDDWIYDLGKYSFT 346
DB 361 TLYLAPGGDDWIYDLGKYSFT 383

RESULT 5

US-08-869-057-2
; Sequence 2, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA

; ZIP: 94804-0099
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: Washtien, Wendy L

; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1

; TELEPHONE: 510-262-5411
; TELEFAX: 510-262-7095

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 423 amino acids
; TYPE: amino acid
; STRANDEDNESS:

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:

; TISSUE TYPE: Plasma
; FEATURE:

; NAME/KEY: Peptide
; LOCATION: 23..401
US-08-869-057-2

Query Match 94.2%; Score 1799.5; DB 2; Length 423;
Best Local Similarity 89.6%; Pred. No. 9.7e-187;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVILFCEQHVFAFOSGOVLAALPRTSRQVQLQNLTTTYEIVLWQPTAD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFOSGOVLAALPRTSRQVQLQNLTTTYEIVLWQPTAD 60

Qy 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVDELIOQISNDTVSPRASASY 120
Db 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVDELIOQISNDTVSPRASASY 120
Qy 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEKYPLYVLKVGKQAKNAIWDGCI 180
Db 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEKYPLYVLKVGKQAKNAIWDGCI 180
Qy 181 HAREWISPAFLWIGH-----NRWRK 203
Db 181 HAREWISPAFLWIGH-----NRWRK 203
Qy 204 NRSFYANNHCLGTDLNSFVSKHCEGASSSSCSETYCGLYPSEPEVKAVAFLLRNI 263
Db 241 NRSFYANNHCLGTDLNSFVSKHCEGASSSSCSETYCGLYPSEPEVKAVAFLLRNI 300
Qy 264 NOIKAYISMHSYSQHIVFPYSTRSKSDHELSLVASEAVRAIDKTSKNTRYTHGHSE 323
Db 301 NOIKAYISMHSYSQHIVFPYSTRSKSDHELSLVASEAVRAIDKTSKNTRYTHGHSE 360
Qy 324 TLYLAPGGDDWIYDLGIKYSFT 346
Db 361 TLYLAPGGDDWIYDLGIKYSFT 383

RESULT 6

US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CLO01173
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-4

Query Match 93.6%; Score 1788.5; DB 4; Length 423;
Best Local Similarity 89.0%; Pred. No. 1.5e-185;
Matches 341; Conservative 1; Mismatches 4; Indels 37; Gaps 1;

Qy 1 MKLCSLAVLPVILFCEQHVAFOSGOVLAALPRTSRQVQLQNTTYYEIVLWQPVTA 60
Db 1 MKLCSLAVLPVILFCEQHVAFOSGOVLAALPRTSRQVQLQNTTYYEIVLWQPVTA 60
Qy 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVDELIOQISNDTVSPRASASY 120
Db 61 LIVKKQVHFFVNASDVNDVKAHLNVSGIPCSVLLADVDELIOQISNDTVSPRASASY 120
Qy 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEKYPLYVLKVGKQAKNAIWDGCI 180
Db 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEKYPLYVLKVGKQAKNAIWDGCI 180
Qy 181 HAREWISPAFLWIGH-----NRWRK 203
Db 181 HAREWISPAFLWIGH-----NRWRK 203
Qy 204 NRSFYANNHCLGTDLNSFVSKHCEGASSSSCSETYCGLYPSEPEVKAVAFLLRNI 263
Db 241 NRSFYANNHCLGTDLNSFVSKHCEGASSSSCSETYCGLYPSEPEVKAVAFLLRNI 300
Qy 264 NOIKAYISMHSYSQHIVFPYSTRSKSDHELSLVASEAVRAIDKTSKNTRYTHGHSE 323
Db 301 NOIKAYISMHSYSQHIVFPYSTRSKSDHELSLVASEAVRAIDKTSKNTRYTHGHSE 360

Qy 324 TLYLAPGGDDWIYDLGIKYSFT 346
Db 361 TLYLAPGGDDWIYDLGIKYSFT 383

RESULT 7

US-08-696-139-2
; Sequence 2, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Payerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Herabberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/696,139
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-696-139-2

Query Match 33.5%; Score 640.5; DB 1; Length 404;
Best Local Similarity 37.7%; Pred. No. 7.6e-61;
Matches 136; Conservative 60; Mismatches 122; Indels 43; Gaps 6;

Qy 23 FQSQVLAALPRTSRQVQLQNTTYYEIVLWQPVTAADLIYKKKQVHFFVNASDVNDVKA 82
Db 10 FEGEKVFRVNVEDENDISLLHELASTRQIDFWKPDSTQIKPHSTVDFRVKAEIDLAV 69
Qy 83 HLNVSIGIPCSVLLADVDELIOQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 142
Db 70 FLEQNELQYEVLLNNLSVLEAQPDSTV---RTTGHSEYKYNWETTEAWTKQVTSNP 126
Qy 143 MLTKIHIGSSFKEKYPLYVLKVGKQAKNAIWDGCIHAREWISPAFLWFI----- 195
Db 127 LISRTALGTTFLGNIIYLLKY-GKPGNKPAIFMDCGFHAREWISHAFQVFEAVLTY 185
Qy 196 -----GH-----NRWRKNSFYANNHCLGTDLNSFVSK 225
Db 186 GYESHMTFELNKLDFYVLPVNLIDGIYTTWTKNRMKTRSTNAGTTTCIGTDPNRNF-DA 244

Best Local Similarity 35.6%; Pred. No. 1.2e-60;
Matches 135; Conservative 69; Mismatches 131; Indels 44; Gaps 7;
QY 6 LALVLPVILFCEQHVFA-FQSGQVLAALPRTSRQVQLNLTITTYEIVLWOPVTADLVK 64
Db 2 LLVLVTVALASAHGGEHFEKGVPRVNVVDENHINIIRLASTTQIDFWKPDSTVTKIP 61
QY 65 KKQVHFFVNASVDVNVKAHLNVSGIPCSVLADVEDLIQQOISNDTVSPRASASYEQYH 124
Db 62 HSTVDFRKAEDTVTVENLVKQELVLSNLRNVVEAQFDSRV---RATGHSYKYN 118
QY 125 SLNEIYSWIFITERHPDMLTKIHIGSFEPKYPVLYLVKSGKEQAKNAIWIDCGIHARE 184
Db 119 KWETIEAWTQVATENPALISRSVIGTFEGRAIYLLKV-SKAGQNKPAIFMDCGFHARE 177
QY 185 WISPAFLWFI-----GH-----NRMWRKNSF 207
Db 178 WISPAFCWFVREAVRTYGREIQVTELDKLDFFVLPVNLIDGYIYTWKSRFWRKTRST 237
QY 208 YANNHCICGLDLSNFVSKHWCCEGASSSCSETYCGLYPSEPEVKAVASFLRNINQIK 267
Db 238 HTGSCIGTDNRF-DAGWEIGASRNPCEYTCGPAASEKETKALADFIKLSIK 296
QY 268 AYISMHSYQHVFPYSTRSKDHELSLVASEAVRAIDKTSKNTRYTHGSGSETLYL 327
Db 297 AYLTHSYQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGKTYTGPATTIYP 355
QY 328 APGGDDWIYDLGIKYSFT 346
Db 356 AAGGDDWAYDQGIKYSFT 374
RESULT 10
US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011,769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-011-769A-56
Query Match 33.0%; Score 631.5; DB 4; Length 424;
Best Local Similarity 35.7%; Pred. No. 7.9e-60;
Matches 129; Conservative 68; Mismatches 121; Indels 43; Gaps 6;
QY 23 FQSGQVLAALPRTSRQVQLNLTITTYEIVLWOPVTADLVKQKQVHFFVNASVDVNVKA 82
Db 29 FEKEKVPVNVVDENHINIIRLASTTQIDFWKPDSTVTKIPHSTVDFRKAEDTVTVEN 88
QY 83 HLNVSIGIPCSVLADVEDLIQQOISNDTVSPRASASYEQYHSLNEIYSWIFITERHPD 142
Db 89 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYKYNKWETIEAWTQVATENPA 145
QY 143 MLTKIHIGSFEPKYPVLYLVKSGKEQAKNAIWIDCGIHAREWISPAFLWFI----- 195
Db 146 LISRSVIGTFEGRAIYLLKV-SKAGQNKPAIFMDCGFHAREWISPAFCWFVREAVRTY 204
QY 196 -----GH-----NRMWRKNSFVANNHCICGLDLSNFVSK 225
Db 205 GREIQVTELDKLDFFVLPVNLIDGYIYTWKSRFWRKTRSTHTGSSCIGTDNRF-DA 263
QY 226 HWCCEGASSSCSETYCGLYPSEPEVKAVASFLRNINQIKAYISMHSYQHVFPYSY 285
Db 264 GWCEIGASRNPCEYTCGPAASEKETKALADFIKLSIKAYLTHSYQMMIYPYSY 323
QY 286 TRSKDHELSLVASEAVRAIDKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 345
Db 324 AYKLGENNAELNALAKATVKEL-ASLHGKTYTGPATTIYPAGGDDWAYDQGIKYSF 382
QY 346 T 346
Db 383 T 383
RESULT 11
US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plaema Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896

TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-649-591B-7

Query Match 32.9%; Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%; Pred. No. 1.6e-59;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;

Qy 6 LAVLVPIVLPCEQHVFQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPVTAIDLIVKK 65
Db 7 MAVIYTTLAIPVH---FDREKVFVRVQLONEKHASVLKNTQSTELDFWYFPAIDHDAVN 63

Qy 66 KOVHFFNVASDVNDVNAHLNVSGIPCSVLLADVEDLIQQOIS-NDTVSPRASASYEYOH 124
Db 64 MTVDVFRVSEKSTQITQSTLEQHKIHYEILHDLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy 125 SLNEIYSWIFETTERHPDMLTKIHIGSSFEKYPYLVKVGSKGQAKNAIWDGCIHARE 184
Db 121 DWDKIVSWTEKMLEKHPMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHARE 179

Qy 185 WISPAFCWFI-----GHN-----RMWRKNRSF 207
Db 180 WISPAFCWFIQATKSYGKNKIMTKLLDRMNFVLPVFNVDGYIWSWTQDRMWRKNRSR 239

Qy 208 YANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPSEPEPVKAVASFLRNINQIK 267
Db 240 NQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTFIRSHLSIK 298

Qy 268 AYISMHSYSQHVPPYSYTRSKSDHELSVASEAVRAIDKTSKNTRYTHGSGSETLYL 327
Db 299 AYITPHSYSQMLLPYGYTFKLPNHDLLKVARIAIDAL-STRYETRYIYGPIASTIYK 357

Qy 328 APGGDDWIDYLGKYSF 345
Db 358 TSGSSLDWYDYGKHTF 375

RESULT 12
US-08-277-540-7
Sequence 7, Application US/08277540
Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Haseak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-7

Query Match 32.9%; Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%; Pred. No. 1.6e-59;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;

Qy 6 LAVLVPIVLPCEQHVFQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPVTAIDLIVKK 65
Db 7 MAVIYTTLAIPVH---FDREKVFVRVQLONEKHASVLKNTQSTELDFWYFPAIDHDAVN 63

Qy 66 KOVHFFNVASDVNDVNAHLNVSGIPCSVLLADVEDLIQQOIS-NDTVSPRASASYEYOH 124
Db 64 MTVDVFRVSEKSTQITQSTLEQHKIHYEILHDLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy 125 SLNEIYSWIFETTERHPDMLTKIHIGSSFEKYPYLVKVGSKGQAKNAIWDGCIHARE 184
Db 121 DWDKIVSWTEKMLEKHPMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHARE 179

Qy 185 WISPAFCWFI-----GHN-----RMWRKNRSF 207
Db 180 WISPAFCWFIQATKSYGKNKIMTKLLDRMNFVLPVFNVDGYIWSWTQDRMWRKNRSR 239

Qy 208 YANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPSEPEPVKAVASFLRNINQIK 267
Db 240 NQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTFIRSHLSIK 298

Qy 268 AYISMHSYSQHVPPYSYTRSKSDHELSVASEAVRAIDKTSKNTRYTHGSGSETLYL 327
Db 299 AYITPHSYSQMLLPYGYTFKLPNHDLLKVARIAIDAL-STRYETRYIYGPIASTIYK 357

Qy 328 APGGDDWIDYLGKYSF 345
Db 358 TSGSSLDWYDYGKHTF 375

RESULT 13
US-08-430-787A-7
Sequence 7, Application US/08430787A
Patent No. 5593674
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995

CLASSIFICATION: 514
PRIOR APPLICATION DATA: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-7

Query Match 32.9%; Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%; Pred. No. 1.6e-59;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;

Qy 6 LAVLVPIVLCFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTVEIVLWQPVTDLIVKK 65
Db 7 MAVIYTTIAIPVH---FDRKQVPRVKLQNEKHASVKNLQTSIELDFWYPDALHDAVN 63

Qy 66 KQVHFFVNASDVNDVNVKAHLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEYQYH 124
Db 64 MTVDVFRVSEKESQITQSLQKHQHYELIHLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy 125 SLNEIYSWIEFITERHPDMLTKIHGSGFEKYPYLVKVSKEQOTAKNAIWDGCIHARE 184
Db 121 DWDKIVSWTEKWLKHPMVSRKIGSTVEDNPLYVLKI-GKDGKRAIPMDGCIHARE 179

Qy 185 WISPAFCWLF-----GHN-----RMRKRSF 207
Db 180 WISPAFCWLFYQATKSYGKQKIMTKLLDRNFFVLPVFNVDGYVWSQDRMRKRSR 239

Qy 208 YANNHCIGTDLNSNFVSKWCEGASSSCSSETYCGLYPESEPEVKAVASFLRRNINQIK 267
Db 240 NQNSTCIGTDLNRF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLSIK 298

Qy 268 AYISMHSYQHVFPYSTRSKSDHBEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
Db 299 AYITFHSYQMLLPYGYTFKLPNHDLLKVARIAITDAL-STRYETRYIYGPIASTIYK 357

Qy 328 APGGDDWYDLGKYSF 345
Db 358 TSGSLDWYDLGIRHTF 375

RESULT 14
US-09-171-945-125
Sequence 125, Application US/09171945
Patent No. 6277599
GENERAL INFORMATION:
APPLICANT: Emery, Stephen
APPLICANT: Copley, Clive Graham
APPLICANT: Edge, Michael Derek
TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
TITLE OF INVENTION: Antibody, and Their Therapeutic Use in an Adept System
FILE REFERENCE: Monoclonal Antibody to CEA
CURRENT APPLICATION NUMBER: US/09171,945
CURRENT FILING DATE: 1998-10-29

PRIOR APPLICATION NUMBER: GB9703103.3
PRIOR FILING DATE: 1997-02-14
PRIOR APPLICATION NUMBER: GB9609405.7
PRIOR FILING DATE: 1996-05-04
PRIOR APPLICATION NUMBER: PCT/GB97/01165
PRIOR FILING DATE: 1997-04-29
NUMBER OF SEQ ID NOS: 131
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 125
LENGTH: 716
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-125

Query Match 32.6%; Score 623; DB 3; Length 716;
Best Local Similarity 34.6%; Pred. No. 1.5e-58;
Matches 132; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

Qy 3 LCSLAVLVPIVLCFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTVEIVLWQPVTDL 61
Db 1 MLALLVLVTVALASAAHGGHFEKEKVRVNVEDENHNIRIELASTTQIDFWKPDVSTQ 60

Qy 62 IVKKQVHFFVNASDVNDVNVKAHLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYE 121
Db 61 IKHSTVDVFRVKAEDTIVENVLQNELQYKVLISLNRNVVEAQFDSRV---RATGHSYE 117

Qy 122 QYHSLNEIYSWIEFITERHPDMLTKIHGSGFEKYPYLVKVSKEQOTAKNAIWDGCIH 181
Db 118 KYNKWTETIAWTQOVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDGCFH 176

Qy 182 AREWISPAFCWLF-----GH-----NRMWRKN 204
Db 177 AREWISPAFCWLFVREAVRTYGREIQVTELLDKLDFVLPVNLIDGYIYTWTSRFRWKT 236

Qy 205 RSFYANNHCIGTDLNSNFVSKWCEGASSSCSSETYCGLYPESEPEVKAVASFLRRNIN 264
Db 237 RSTHTGSCIGTDLNRF-DAGWCEIGASRNPCDETYCGPAASEKETKALADPIRNKLS 295

Qy 265 QIKAYISMHSYQHVFPYSTRSKSDHBEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
Db 296 SIRAYLTHSYQMWIYPYVAYKLGENNAELNALAKATVKEL-ASLHGTYKYTYGPGATT 354

Qy 325 LYLAPGGDDWYDLGKYSFT 346
Db 355 IYPSAGTSKDWAYDQGIYSFT 376

RESULT 15
US-07-649-591B-6
Sequence 6, Application US/07649591B
Patent No. 5206161
GENERAL INFORMATION:
APPLICANT: Dennis Drayna and Daniel Eaton
TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/649,591B
FILING DATE: 19910201
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/266-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-649-591B-6

Query Match 32.4%; Score 619.5; DB 1; Length 417;
Best Local Similarity 34.1%; Pred. No. 1.5e-58;
Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;
Qy 6 LALVPIVLFCQEHVEA---FQSQVLAAALPRTSRQVQLQLTTTVEIVLWQPVTDLI 62
Db 1 MRLILPVLGIATTLAIAPVRPDRKVRVQDEKQADIINDLAKTNELDFWYPGATHV 60
Qy 63 VKKKQVHFFVNASVDVNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYE 121
Db 61 AANMVDVFRVSEKESQAIQSALDQNMKHYELIHDLOEEIEKQFDVKEDIPGRHS---YA 117
Qy 122 QYHSLNFIYSWIEFITERHPDMLTKIHHGSSFEXKYPLYVLKVSKEQTAKNAIWDGCIH 181
Db 118 KYNWKEIVAWTEKOMDKPEMVSRIKIGSTVEDNPPLYVLKI-GEKNERRKAIIFMDCGIH 176
Qy 182 AREWISPAFLWFI-----GH-----NRMWRKN 204
Db 177 AREWSPAFQWFYQATKYGRNKIMTKLLDRNFVILPVFNVDGYIWSWKRMWRKN 236
Qy 205 RSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
Db 237 RSKNQSKCIGTDLNRF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLN 295
Qy 265 QIKAYISMHSYSHIVFPYSTRSKSDHELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
Db 296 EIKVYITFHSYQMLFPYGYTSKLPNNHEDLAKVAKIGTDVL-STRYTRYIYGPIEST 354
Qy 325 LYLAPGGDDWIYDLGIKYSF 345
Db 355 IYPISSGSLDWAYDLGIRHTF 375

Search completed: January 1, 2004, 19:33:09
Job time : 24.1712 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:32:17 ; Search time 51.4823 Seconds
(without alignments)
1403.024 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLPVILFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1799.5	94.2	423	12	US-10-379-836-17
2	1716.5	89.8	423	12	US-10-379-836-2
3	1528	80.0	422	12	US-10-379-836-18
4	1499	78.4	422	12	US-10-379-836-16
5	803.5	42.0	211	9	US-09-925-302-467
6	631.5	33.0	402	12	US-10-379-836-20
7	623	32.6	716	12	US-09-910-059-125
8	619.5	32.4	417	12	US-10-341-434-188
9	619.5	32.4	417	15	US-10-229-546-2
10	619.5	32.4	417	15	US-10-229-546-9
11	604	31.6	437	12	US-10-274-639-12
12	604	31.6	437	14	US-10-200-344-10
13	563	29.5	374	9	US-09-888-615-61
14	563	29.5	444	16	US-10-176-306-74
15	546	28.6	399	15	US-10-200-910-8

16	543	28.4	613	10	US-09-910-059-113	Sequence 113, App
17	528.5	27.7	350	14	US-10-200-344-12	Sequence 12, Appl
18	519.5	27.2	436	15	US-10-200-910-6	Sequence 6, Appl
19	511.5	26.8	436	14	US-10-076-535-2	Sequence 2, Appl
20	508	26.6	417	9	US-09-923-779-150	Sequence 150, App
21	504	26.4	428	9	US-09-925-297-528	Sequence 528, App
22	495.5	25.9	298	12	US-10-379-836-19	Sequence 19, Appl
23	489.5	25.6	406	9	US-09-925-297-517	Sequence 517, App
24	489.5	25.6	419	12	US-10-345-680-65	Sequence 65, Appl
25	483	25.3	421	11	US-09-946-374-234	Sequence 234, App
26	483	25.3	421	12	US-10-015-387A-234	Sequence 234, App
27	483	25.3	421	12	US-10-006-130A-234	Sequence 234, App
28	483	25.3	421	12	US-10-199-672-308	Sequence 308, App
29	483	25.3	421	12	US-10-006-172A-234	Sequence 234, App
30	483	25.3	421	12	US-10-187-749-308	Sequence 308, App
31	483	25.3	421	12	US-10-194-457-308	Sequence 308, App
32	483	25.3	421	12	US-10-184-643-308	Sequence 308, App
33	483	25.3	421	12	US-10-196-747-308	Sequence 308, App
34	483	25.3	421	12	US-10-015-392A-234	Sequence 234, App
35	483	25.3	421	12	US-10-017-253A-234	Sequence 234, App
36	483	25.3	421	12	US-10-173-689-308	Sequence 308, App
37	483	25.3	421	12	US-10-173-690-308	Sequence 308, App
38	483	25.3	421	12	US-10-173-691-308	Sequence 308, App
39	483	25.3	421	12	US-10-173-692-308	Sequence 308, App
40	483	25.3	421	12	US-10-173-694-308	Sequence 308, App
41	483	25.3	421	12	US-10-173-698-308	Sequence 308, App
42	483	25.3	421	12	US-10-173-699-308	Sequence 308, App
43	483	25.3	421	12	US-10-173-707-308	Sequence 308, App
44	483	25.3	421	12	US-10-174-569-308	Sequence 308, App
45	483	25.3	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT FILING DATE: 2003-03-04
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Query Match	94.2%	Score 1799.5;	DB 12;	Length 423;
Best Local Similarity	89.6%	Pred. No. 2.9e-175;		
Mismatches	343;	Conservative	1;	Mismatches 2;
Indels	37;	Gaps	1;	
Qy	1	MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQVQLQNTTYYEIVLWQPVTD	60	
Db	1	MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQVQLQNTTYYEIVLWQPVTD	60	
Qy	61	LIYKKQVHFVNASVDNKAHLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASYY	120	
Db	61	LIYKKQVHFVNASVDNKAHLNVSIGPCSVLLADVEDLIQOOISNDTVSPRASYY	120	
Qy	121	EQVHSLNEIYSWIEFTITERHPDMLTKIHIGSSPEKYPYLYLVKSGKEQTAKNAIWDGCI	180	
Db	121	EQVHSLNEIYSWIEFTITERHPDMLTKIHIGSSPEKYPYLYLVKSGKEQTAKNAIWDGCI	180	
Qy	181	HAREWISPAFCLWFIGH-----NRMRK	203	

Db 181 HAREWISPAFLWFIHTQFYGIQYTNLLRLVDFVMPVNVVDGYDYSWKKNRWRK 240
Qy 204 NRSFYANNHCITGLNSNFVSKHCEGASSSCSETYCGLYPSEPEVKAVASFLRNI 263
Db 241 NRSFYANNHCITGLNRRNFASKHCEGASSSCSETYCGLYPSEPEVKAVASFLRNI 300
Qy 264 NQIKAYISMHSYSOHIVPPYSTRSKSKDHHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
Db 301 NQIKAYISMHSYSOHIVPPYSTRSKSKDHHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 360
Qy 324 TLYLAPGGDDWIYDLGKIYKFT 346
Db 361 TLYLAPGGDDWIYDLGKIYKFT 383

RESULT 2

US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Papio hamadryas
US-10-379-836-2

Query Match 89.8%; Score 1716.5; DB 12; Length 423;
Best Local Similarity 85.4%; Pred. No. 9e-167;
Matches 327; Conservative 7; Mismatches 12; Indels 37; Gaps 1;

Qy 1 MKLCSLAVLPIVLFCEQHVAFQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTA 60
Db 1 MKLCSLAVLPIVLFCEQHVAFQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTA 60
Qy 61 LIVKKQVHFVNADVDNVKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASAY 120
Db 61 LIEKKQVHFVNSSVDNVKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASAY 120
Qy 121 EQVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKSGKQTKAKNAIWD 180
Db 121 EQVHSLNEIYSWIELETKYDPMLETKIHGSSYKHPYLYLVKSGKQTKAKNAIWD 180
Qy 181 HAREWISPAFLWFIHTQFYGIQYTNLLRLVDFVMPVNVVDGYDYSWKKNRWRK 203
Db 181 HAREWISPAFLWFIHTTEYGIIEYTNLLRHVDVYMPVNVVDGYDYSWKKNRWRK 240
Qy 204 NRSFYANNHCITGLNSNFVSKHCEGASSSCSETYCGLYPSEPEVKAVASFLRNI 263
Db 241 NRSFYANNHCITGLNRRNFASKHCEGASSSCSETYCGLYPSEPEVKAVASFLRNI 300
Qy 264 NQIKAYISMHSYSOHIVPPYSTRSKSKDHHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
Db 301 NQIKAYISMHSYSOHIVPPYSTRSKSKDHHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 360
Qy 324 TLYLAPGGDDWIYDLGKIYKFT 346
Db 361 TLYLAPGGDDWIYDLGKIYKFT 383

RESULT 3

US-10-379-836-18
; Sequence 18, Application US/10379836

; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-379-836-18

Query Match 80.0%; Score 1528; DB 12; Length 422;
Best Local Similarity 76.2%; Pred. No. 1.8e-147;
Matches 292; Conservative 24; Mismatches 29; Indels 38; Gaps 2;

Qy 1 MKLCSLAVLPIVLFCEQHVAFQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTA 60
Db 1 MKLHGLGILVAITILY-EQHGFAFGQGVLSALPRTSRQVQLNLTYYEIVLWQPVTA 59
Qy 61 LIVKKQVHFVNADVDNVKAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASAY 120
Db 60 FIEKKQVHFVNADVDNVKAHLNVSRIPFVLMNVEDLIEQOITFNDTVSPRASAY 119
Qy 121 EQVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKSGKQTKAKNAIWD 180
Db 120 EQVHSLNEIYSWIEVITEQHPDMLQKIYIGSSPEKPYLYLVKSGKQTKAKNAIWD 179
Qy 181 HAREWISPAFLWFIHTQFYGIQYTNLLRLVDFVMPVNVVDGYDYSWKKNRWRK 203
Db 180 HAREWISPAFLWFIHTQFYGHGKENVLYTRLLRHVDVYMPVNVVDGYDYSWKKNRWRK 239
Qy 204 NRSFYANNHCITGLNSNFVSKHCEGASSSCSETYCGLYPSEPEVKAVASFLRNI 263
Db 240 NRSFYANNHCITGLNRRNFASKHCEGASSSCSETYCGLYPSEPEVKAVADFLRNI 299
Qy 264 NQIKAYISMHSYSOHIVPPYSTRSKSKDHHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
Db 300 DHKAYISMHSYSOHIVPPYSTRSKSKDHHEELSLVASEAVRAIESINKNTRYTHGHGSE 359
Qy 324 TLYLAPGGDDWIYDLGKIYKFT 346
Db 360 SLYLAPGGDDWIYDLGKIYKFT 382

RESULT 4

US-10-379-836-16
; Sequence 16, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-379-836-16

Query Match 78.4%; Score 1499; DB 12; Length 422;

PRIOR APPLICATION NUMBER: 60/256,249
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: 60/256,405
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: 10/083,248
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: PCT/US01/46717
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/242,324
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/242,518
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/241,989
PRIOR FILING DATE: 2000-10-20
NUMBER OF SEQ ID NOS: 86
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 74
LENGTH: 444
TYPE: PRT
ORGANISM: Homo sapiens
US-10-176-306-74

Query Match 29.5%; Score 563; DB 16; Length 444;
Best Local Similarity 38.7%; Pred. No. 1.1e-48;
Matches 109; Conservative 51; Mismatches 82; Indels 40; Gaps 5;

Qy 103 QOQISNDTVSPRASASY-YEOYHSLNEIYSWIEFITERHPDMLTKIHIGSFKEKYPYVL 161
Db 159 ROEIVDKSVSPWLETYSYNIHPMGEIYEMWREISEKYKEVVTQHFLGVTYETHPIYVL 218
Qy 162 KVSQEQTAKNAIWDGCIHAREWISPAFLWFI----- 195
Db 219 KISQPSGNPKKIIWMDGCIHAREWIAPAFQCMFKVEILQNHKNSRIRKLRNLDIFYVLP 278
Qy 196 -----GH-----NRWRKNRSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSETYCGL 244
Db 279 VLNDIGYIYTTTDLRWKRSPPNNGTCFCTDLNRNF-NASWCSIGASRNCQDTFCGT 337
Qy 245 YPESEPEKAVASFLRNINOIKAYISMHSYQHVFPYSYTRSKSDHELSLVASEAV 304
Db 338 GPVSEPETKAVASFIESKKDDILCFLTWHSYGQLILTPYGTKNKSNHPMIQVGQKAA 397
Qy 305 RAIDKTSKNTRYTHGHSETLYLAPGGDDWIYDLGIKYSFT 346
Db 398 NAL-KAKYGTNYRVGSSADILYASSGSSRDRWARDIGIPFSYT 438

RESULT 15

US-10-200-910-8
Sequence 8, Application US/10200910
Publication No. US20030009021A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Nehls, Michael
APPLICANT: Friedrich, Glenn
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
TITLE OF INVENTION: No. US20030009021A1el Human Proteases and
FILE REFERENCE: LEX-0086-USA
CURRENT APPLICATION NUMBER: US/10/200,910
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US/09/710,099
PRIOR FILING DATE: 2000-11-10
PRIOR APPLICATION NUMBER: US 60/165,260
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 399
TYPE: PRT

ORGANISM: homo sapiens
US-10-200-910-8

Query Match 28.6%; Score 546; DB 15; Length 399;
Best Local Similarity 38.7%; Pred. No. 5.3e-47;
Matches 129; Conservative 60; Mismatches 124; Indels 20; Gaps 10;

Qy 23 FQSGQVLAALPRTSRQVQVLQNL--TTYEIVLWQ--PVTADLIYKKQVHFFVNASVDVN 79
Db 37 FTGDQVLRVLAKDEKQLSLGLLEGKPKQKVDWFRGPARPSLPVDMR-----VPFSLKD 91
Qy 80 VKAHLNVSGIPCSVLLADVEDLI--QQQISNDTVSPRASASY-YEOYHSLNEIYSWIEF 135
Db 92 IKAYLESHGLAYSIMIKDIQVLLDEERQAWAKSRRLERSTNSFSYSYHTLEIYSIDN 151
Qy 136 ITERHPDMLTKIHIGSFKEKYPYVLKVS--GKEQTAKNAIWDGCIHAREWISPAFLW 193
Db 152 FVMEHSDIVSKIQIGNSFENQSIILVKLFSTGGRHP--AIWIDTGIHSREWIHTATGIW 208
Qy 194 FIGHNRWRKNRSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSETYCGLYPESEPEVK 253
Db 209 TANKNRLWRKNKSGIRPGIFCIGVDLRNWK-S-GFGNGSNSNFCSETYHGPSPQSEPEVA 267
Qy 254 AVASEFLRNINQIKAYISMHSYQHVFPYSYTRSKSKDHELSLVASEAVRAIDKTSKN 313
Db 268 AIVNFITAHGN-FKALISIIHSYQMLPYGRLLLEPVSNORELYDLAKDAVEALYKV-HG 325
Qy 314 TRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346
Db 326 IEYIFGSISTTLYVASGITVDMAYDSGICKYAFS 358

Search completed: January 1, 2004, 19:50:56
Job time : 52.4823 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:25:56 ; Search time 24.8017 Seconds
(without alignments)
1395.903 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1799.5	94.2	423	2 A41204	carboxypeptidase B
2	692	36.2	415	2 A32129	carboxypeptidase B
3	628.5	32.9	417	1 A34487	carboxypeptidase A
4	621.5	32.5	416	1 A42332	carboxypeptidase B
5	619.5	32.4	417	2 A43929	carboxypeptidase B
6	580	30.4	306	1 CPBOB	carboxypeptidase A
7	553	28.9	417	2 A32128	carboxypeptidase A
8	546	28.6	309	2 A38395	mast cell carboxyp
9	533.5	27.9	419	1 CPBOA	carboxypeptidase A
10	528.5	27.7	419	1 CPRTA	carboxypeptidase A
11	508	26.6	417	2 A56171	carboxypeptidase A
12	489.5	25.6	419	1 S29127	carboxypeptidase A
13	437.5	22.9	528	2 T33527	hypothetical prote
14	397	20.8	323	2 T20507	hypothetical prote
15	395.5	20.7	667	2 T33526	hypothetical prote
16	391.5	20.5	527	2 T27572	hypothetical prote
17	369.5	19.3	303	1 CPCYB	carboxypeptidase B
18	362.5	19.0	455	2 T33572	hypothetical prote
19	358	18.7	666	2 T24170	hypothetical prote
20	299.5	15.7	430	2 S48976	ECM14 protein - ye
21	277.5	14.5	424	2 S17571	carboxypeptidase T
22	271	14.2	451	2 S20723	carboxypeptidase (
23	250	13.1	497	2 T40260	carboxypeptidase p
24	244.5	12.8	373	2 T26030	hypothetical prote
25	173	9.1	999	2 T36021	probable zinc-bind
26	135.5	7.1	1446	2 T30916	carboxypeptidase D
27	130.5	6.8	558	2 S75104	hypothetical prote
28	118.5	6.2	707	2 T40070	origin recognition
29	104.5	5.5	985	2 D82776	pyruvate dehydroge

30	103	5.4	1216	2 H84629	hypothetical prote
31	101.5	5.3	985	2 T29910	hypothetical prote
32	100	5.2	376	1 B69957	gamma-D-glutamyl-L
33	99.5	5.2	1389	2 I50090	carboxypeptidase g
34	99	5.2	1587	2 A82012	hypothetical prote
35	98.5	5.2	983	2 B49284	immediate-early pr
36	98.5	5.2	1078	2 T44232	hypothetical prote
37	98	5.1	491	2 B96739	hypothetical prote
38	97	5.1	663	2 T37772	telomere length re
39	96	5.0	368	2 T21748	hypothetical prote
40	95.5	5.0	355	2 C83850	gamma-D-glutamyl-L
41	95	5.0	444	2 S51199	cryptophan 5-monoo
42	95	5.0	662	2 G86210	hypothetical prote
43	94	4.9	606	2 E90536	hypothetical prote
44	93.5	4.9	986	2 E90596	restriction-modifi
45	93.5	4.9	986	2 H90565	restriction modifi

ALIGNMENTS

RESULT 1
A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N:Alternate names: plasma carboxypeptidase B
C:Species: Homo sapiens (man)
C:Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C:Accession: A41204
R:Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A:Title: Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase
A:Reference number: A41204; MUID:92042093; PMID:1939207
A:Accession: A41204
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-423 <EAT>
A:Cross-references: GB:M75106; NID:G189686; PID:G189687
C:Genetics:
A:Gene: GDB:CPB2
A:Cross-references: GDB:129546; OMIM:212070
A:Map position: 13q14.11-13q14.11
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-114/Domain: activation peptide #status predicted <ACP>
F:115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Query Match 94.2%; Score 1799.5; DB 2; Length 423;
Best Local Similarity 89.6%; Pred. No. 3.2e-140;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

Qy	1	MKLCSLAVLVPIVLFCEQHVFAFGQVLAALPRTSRQVQLQNLTTTTEIVLMQPVTTAD 60
Db	1	MKLCSLAVLVPIVLFCEQHVFAFGQVLAALPRTSRQVQLQNLTTTTEIVLMQPVTTAD 60
Qy	61	LIVKKQVHFVNASDVNDVNVKAHLNVSGIPCSVLADVEDLIQQIISNDTVSPRASASY 120
Db	61	LIVKKQVHFVNASDVNDVNVKAHLNVSGIPCSVLADVEDLIQQIISNDTVSPRASASY 120
Qy	121	EQYHSLNIYSWISFITERHPDMLTKIHIGSPFKYPLYVLKVSKEQTAKNAIWDGCI 180
Db	121	EQYHSLNIYSWISFITERHPDMLTKIHIGSPFKYPLYVLKVSKEQTAKNAIWDGCI 180
Qy	181	HAREWISPAFCFLWFTGH-----NRWNRK 203
Db	181	HAREWISPAFCFLWFTGHITQFYIGIQTNLRLVDFYVPPVNVVDGYDYSKKNRMRK 240
Qy	204	NRSFYANNHCIGTDLNSNFVSKWCCEGASSSSCSSETYCGLYPSEPEPKAVASFLRRNI 263
Db	241	NRSFYANNHCIGTDLNRNFASKWCCEGASSSSCSSETYCGLYPSEPEPKAVASFLRRNI 300
Qy	264	NOIKAYISMHSYQHIVFPYYSYTRSKSKDHELSLVAEAVRAIDKTSKNTRYTHGHGSE 323

Db 301 NQIKAYISMHSYQHIVPYPYSYTRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGHSE 360
QY 324 TLYLAPGGDDWIYDLGIKYSFT 346
Db 361 TLYLAPGGDDWIYDLGIKYSFT 383

RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C:Accession: A32129; S17543
R:Clauer, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A>Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparison of the two genes.
A:Reference number: A32693; MUID:89034324; PMID:3182872
A:Accession: A32129
A:Molecule type: DNA
A:Residues: 1-415 <CLA>
A:Cross-references: GB:M23959; GB:J04041; NID:G203293; PIDN:AAA40872.1; PID:G203295
R:Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A>Title: Distribution of manganese in rat pancreas and identification of its primary binding site.
A:Reference number: S17543; MUID:91378950; PMID:1898371
A:Accession: S17543
A:Molecule type: protein
A:Residues: 109-116, 'X', 118-130 <KOD>
C:Genetics: CPB
A:Gene: CPB
A:Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:354,376/Active site: Tyr, Glu #status predicted

Query Match 36.2%; Score 692; DB 2; Length 415;
Best Local Similarity 39.2%; Pred. No. 4e-49;
Matches 150; Conservative 61; Mismatches 126; Indels 46; Gaps 7;

QY 1 MKLCSLAVLPVILFCEQHVAFQSGQVLAALPRTSRQVQLQNLTTTVEIVLWQPVAD 60
Db 1 MLLLLALVSLAHASEEH---FDGNRVYRVSVHGEDHVNLIQELANTKEIDFWKPD SAT 57

QY 61 LIVKKQVHFFYNASDVNDVNAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASASY 120
Db 58 QVKKPTTVDHFVKAEDVADVENFLEENVEHYEVLISNRVALESQFDSHT---RASGHSY 114

QY 121 EGYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVSKEQTAKNAIWDGCI 180
Db 115 TYKNWETLEAIQVATDNDPLVTQSVIGTTFEGRNMYVLKI-GKTRPKPAIFIDCGF 173

QY 181 HAREWISPAFLWFI-----GH-----NRMWRK 203
Db 174 HAREWISPAFCQWFVREAVRTYQBIHMKQLLDLDFVYLPVNVNIDGVYVYTWTKDRMWRK 233

QY 204 NRSFYANNHCIGTDLNSNFVSKHWCEGASSSSCSETYCGLYPESEPEKAVASFLRNLI 263
Db 234 TKTSTWAGSCLGVRRNRP-NAGWCVEVGASRSPCSETYCGPAPESEKETKALADFI RNLL 292

QY 264 NQIKAYISMHSYQHIVPYPYSYTRSKSKDHELSLVASEAVRAIDKTSKNTRYTHGHSE 323
Db 293 STIKAYLTHISYQMLYPYSYDYKLPENYELNALVGAKEI-ATLHGTYTYGPGAT 351

QY 324 TLYLAPGGDDWIYDLGIKYSFT 346
Db 352 TIYPAAAGGDDWSYQGIKYSFT 374

RESULT 3
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A34487
R:Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A>Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member of the carboxypeptidase A family.
A:Reference number: A34487; MUID:90062123; PMID:2584208
A:Accession: A34487
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-417 <REY>
A:Cross-references: GB:J05118; NID:G192372; PIDN:AAA37369.1; PID:G309135
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:356,378/Active site: Tyr, Glu #status predicted

Query Match 32.9%; Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%; Pred. No. 6.8e-44;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;

QY 6 LAVLVPVILFCEQHVAFQSGQVLAALPRTSRQVQLQNLTTTVEIVLWQPVADLIVKK 65
Db 7 MAVYTTLAIAPVH---FDRKVFVKLVKQNEKHAASVKNLTQSIELDFWYDAIHDAVN 63

QY 66 KQVHFFYNASDVNDVNAHLNVSGIPCSVLLADVEDLIQQOIS-NDTVSPRASASYEQYH 124
Db 64 MTVDPRVSEKESQITQSTLEQHKIHYELIHDLOBEIEKQFDVKDEIAGRHS---YAKYN 120

QY 125 SLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVSKEQTAKNAIWDGCIHARE 184
Db 121 DWDKIVSWTERKMLEKHPMVSRIKIGSTVEDNPPLYVLKI-GKKDGERKAI FMDCCGIHARE 179

QY 185 WISPAFLWFI-----GH-----RMWRKNSRF 207
Db 180 WISPAFCQWFYQATKSYGKNKIMTKLLDRNFFVLPVFNVDGYIWSQTDRMWRKNSR 239

QY 208 YANNHCIGTDLNSNFVSKHWCEGASSSSCSETYCGLYPESEPEKAVASFLRNINQIK 267
Db 240 NQNSTCIGTDLNRF-DVSWDSSPNTKPCLVNVRGPAPASEKETKATVNFIRSHLSIK 298

QY 268 AYISMHSYQHIVPYPYSYTRSKSKDHELSLVASEAVRAIDKTSKNTRYTHGHSETLYL 327
Db 299 AYITFHSYQMLLPYGYTFKLPPNHQDLKVARIATDAL-STRYETRYIYGPIASTIYK 357

QY 328 APGGDDWIYDLGIKYSF 345
Db 358 TSGSLDWDYDLGIKHTF 375

RESULT 4
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N:Alternate names: carboxypeptidase B isozyme 1
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C:Accession: A42332; S08254; S02813; S02812
R:Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Sham, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992
A>Title: Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of the cDNA.
A:Reference number: A42332; MUID:92129345; PMID:1370825
A:Accession: A42332
A:Molecule type: mRNA
A:Residues: 1-416 <YAM>
A:Cross-references: GB:M81057; NID:G809194; PIDN:AAA66973.1; PID:G189625
A:Experimental source: pancreas
A>Note: sequence extracted from NCBI backbone (NCBI:78573, NCBIP:78574)
R:Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A>Title: Further studies on the human pancreatic binary complexes involving procarboxypeptidase B.
A:Reference number: S08253; MUID:90169111; PMID:2307232
A:Accession: S08254
A:Molecule type: protein
A:Residues: 16-40 <MOU>
A:Experimental source: pancreas
R:Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.

230 LKLVIIIRGSGWMEFFGIGAGKGFEMEDLVNVAAGSDVLE SIKRIRIKTIGFADBI 334
 QY 325 LYLAPGGGDDWIYDLGIKYSF 345
 : | | | | | : |
 : | | | | | : |
 Db 355 IYPISGSLDWAYDLGIKHTF 375

 RESULT 6
 CPBOB
 C:carboxypeptidase B (EC 3.4.17.2) - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 24-Apr-1984 #sequence revision 24-Apr-1984 #text_change 02-Aug-1994
 C:Accession: A93797; A92150; A00912
 R:Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
 Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
 A:Title: Amino-acid sequence of bovine carboxypeptidase B.
 A:Reference number: A93797; MUID:75217824; PMID:1057162

A:Accession: A93797
A:Molecule type: protein
A:Residues: 1-306 <TIR>
R:Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A:Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations of
A:Reference number: A92150; MUID:74260705; PMID:4833744
A:Accession: A92150
A:Molecule type: protein
A:Residues: 31-93; 131-181; 263-265; 292-306 <SCH>
R:Schmidt, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A:Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A:Reference number: A92839; MUID:76265065; PMID:957425
A:Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R:Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A:Title: Isolation and sequence of peptides at the active center of bovine carboxypeptidase
A:Reference number: A92051; MUID:70007159; PMID:5344132
A:Contents: annotation; active site
R:Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A:Title: Identification of a glutamic acid at the active center of bovine carboxypeptidase
A:Reference number: A92119; MUID:73061487; PMID:4565668
A:Contents: annotation; active site
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:63-76,135-158,149-163/Disulfide bonds: #status experimental
F:66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F:246,268/Active site: Tyr, Glu #status experimental

Query Match 30.4%; Score 580; DB 1; Length 306;
Best Local Similarity 43.6%; Pred. No. 4.3e-40;
Matches 115; Conservative 44; Mismatches 65; Indels 40; Gaps 5;

QY 120 YQYHSLNEIYSWIFETERHPDMLTKIHIGSSFYKPYLYLVKVGKEQTAKNAIWDG 179
DB 6 YEKYNWETIEATWTEQVASENFDLISRSAGITFLNGTILYLLKV-GKPGSKNPAVFMDCG 64
QY 180 IHAREWISPAFCMLFI-----GH-----NRMR 202
DB 65 FHAREWISPAFCQWFREAVRTYGREIHMTFLDKLDFVLPVWNIDGIYTWTTNRMR 124
QY 203 KNRSFVANNHCITGLNSNFVSKHCEGASSSSCSEYCYGLYPSSEPEVKAVASFLRN 262
DB 125 KTRSTRAGSCGTDLNRNF-DAGWCISGASNNPCSEYCYGSAESEKESKAVADFIRNH 183
QY 263 INQIKAYISMHSYSHIPEYPSYTRSKSDHEELSLVASEAVRAIDKTSKNTRYTHGHS 322
DB 184 LSSIKAYLTIHYSQWMLPYSDYKLPKNVVELTAKGAVKKL-ASLHGTYSGPGA 242
QY 323 EFLYLPAGGGDDWIYDLGKYSFT 346
DB 243 TTIYPASGGSDWAYDQGIKYSFT 266

RESULT 7
A32128
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
A:Accession: A32128
R:Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rub
J. Biol. Chem. 263, 17828-17836, 1988
A:Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and ev
A:Reference number: A32128; MUID:89034323; PMID:3182871
A:Accession: A32128
A:Molecule type: mRNA
A:Residues: 1-417 <GAR>
A:Cross-references: GB:M2719; GB:M23721; GB:J04043; PIDN:AAA0956.1; PID:9
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:356,378/Active site: Tyr, Glu #status predicted

Query Match 28.9%; Score 553; DB 2; Length 417;
Best Local Similarity 33.5%; Pred. No. 1.1e-37;
Matches 137; Conservative 69; Mismatches 137; Indels 66; Gaps 13;

QY 1 MKCLSLAVLPIVLFCEQHFVAFSGQVLAALPRTSRQVQLNLTTT--YEIVLWQPVT 58
DB 1 MRLTLLAALAGIYICQ---TFVGDQVLEIIPSHHEQIRTLQLAESEHLELDFWKSP 57
QY 59 ADLIVKKQVHPFNASVDNVKHAHLNVSIGPCSVLLADVEDLIQQISN---DTVSPPA 115
DB 58 ---IPGETVHRVPFASIQAVKVFLESQGDYSDIMIEDVQVLLDQEREEMLFNQORRG 113
QY 116 SASVYEQVHSLNEIYSWIFETERHPDMLTKIHIGSSFYKPYLYLVKYS-GKEQTAKNAI 174
DB 114 GNFNFEAVHTLEEIYQEMNLVAENPGLVSKNLGSSFENRPMNVNLFKSTGDD---KPAI 170
QY 175 WIDCGIHAREWISPAFCMLW-----FI-----GH----- 197
DB 171 WLDAGIHAREWVTQATALTANKIASDYGDPAITSLTLNTLIDIFLLPVPNPDGYVFSQTT 230
QY 198 NRMWRKNSFYANNHCITGLNSNFVSKHCEGASSSSCSEYCYGLYPSSEPEVKAVAS 257
DB 231 NRMWRKTRSKSGCGVGVDPNRNW-DANFGGPGCASSSPCSDSVHGPKPNSEVEVKSI 289
QY 258 FLRRNINOIKAYISMHSYSHIPEYPSYTRSKSDHEELSLVASEAVRAIDKTSKNTRY 317
DB 290 FIKSH-GKVPFATILHSYQLMFPYGYKCTKPDDEFNDELDEVAQAAQAL-KRLHGTSYK 347
QY 318 HGHGSETLYLPAGGGDDWIYDLGKYSFTSN-----PPVEKLIP 356
DB 348 VGPICSVIYQASGGSDWAYDLGIKYSFAFELRDTAFYGFLLPAKQILP 396

RESULT 8
A38395
mast cell carboxypeptidase (EC 3.4.17.1) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
A:Accession: A38395; A33118
R:Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A:Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme acti
A:Reference number: A38395; MUID:91105153; PMID:1988052
A:Accession: A38395
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-309 <COL>
R:Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A:Reference number: A33118
A:Accession: A33118
A:Status: preliminary
A:Molecule type: protein
A:Residues: 2-309 <CO2>
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase
F:248,270/Active site: Tyr, Glu #status predicted

Query Match 28.6%; Score 546; DB 2; Length 309;
Best Local Similarity 42.2%; Pred. No. 2.8e-37;
Matches 111; Conservative 39; Mismatches 73; Indels 40; Gaps 5;

QY 120 YEQYHSLNEIYSWIFETERHPDMLTKIHIGSSFYKPYLYLVKVGKEQTAKNAIWDG 179
DB 8 YAKYNDWNKIVSWTEKMKVEKHPMVSRIKIGSTVEDNPLYVLKI-GRKDGKKAIFMDCG 66
QY 180 IHAREWISPAFCMLFI-----GH-----NRMR 202
DB 67 IHAREWISPAFCQWFVYQAAKSYGKNINMTKLLDRNPFVLPVFNVDGYINSWTKDRMR 126
QY 203 KNRSFVANNHCITGLNSNFVSKHCEGASSSSCSEYCYGLYPSSEPEVKAVASFLRN 262
DB 243 TTIYPASGGSDWAYDQGIKYSFT 266

A;Residues: 1-260,'F',262,'M',264-346,'K',348-419 <CLA>

C;Genetics:

A;Gene: CPA1

A;Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1

A;Note: the authors translated the codon GTC for residue 196 as file

C;Superfamily: carboxypeptidase

C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-110/Domain: activation peptide #status predicted <PRO>

F;Fill-419/Product: carboxypeptidase A #status predicted <Mar>

F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted

F;248-271/Dsulfide bonds: #status predicted

F;358,380/Active site: Tyr, Glu #status predicted

Query Match 27.7%; Score 528.5; DB 1; Length 419;

Best Local Similarity 34.4%; Pred. No. 1.2e-35;

Matches 135; Conservative 61; Mismatches 133; Indels 63; Gaps 11;

QY 2 KLCSLAVLPVIFCFQHVAFQSGQVLAALPRTSRQVQLQNLTTTVEIVL--WQPVTA 59

DB 3 RLILSLLEAVCGNEN----FVGHQVLRISAADQAQVKLEDELEHLQDLDFRDAAR 58

QY 60 DLIVKKQVHFVNASDVNDVNAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRAS 116

DB 59 AGPIDVRVPF----PSIQSVKAFLEYHGISEIMIEDVQLLDDEKQMSAFQARALST 114

QY 117 ASY-YEQYHSLNEIYSWIEFITERHDPDLTKIHGSSPEKYPYLVKVGKQETAKNAIW 175

DB 115 DSNFYATYHLDIEYEFMDLLVAEHPQLVSKIQIGNTEGRPIHVLKFS-TGQTNRAIW 173

QY 176 IDCGTHAREWISPAFLWFIHG-----N 198

DB 174 IDTGHSREWVQAGSWFAKKVTKDYGDPTFTAVLNDMDIFLEIVNPDGFATHTKN 233

QY 199 RMRWKNRSFYANNHCIGTDLNSFVSKHW----CEEAGSSSCSETYCGLYPESEPEVKA 254

DB 234 RMRWKNRSHTQGLCVGVDNPN-----WDAGLGKAGASSNFCSETYRGKPNSEVEVKS 288

QY 255 VASFLRNINQIKAYISMHSYQHIVFPYSTRSKSDHEELSLVASEAVRAIDTKSKNT 314

DB 289 IVDFTVSHGN-IKAFISHSYQLLYPYGYTSEPAPDQALDQLAKSAVTAL-TSLHGT 346

QY 315 RYTHGHSSTLYLAPGGDDWYDLGKYSFT 346

DB 347 EFKYGIIDTIYQAGSTIDWTYSOGIKYSFT 378

RESULT 11

A56171

carboxypeptidase A2 (EC 3.4.17.15) precursor - human

N;Alternate names: pancreatic carboxypeptidase A2

C;Species: Homo sapiens (man)

C;Date: 28-Apr-1995 #sequence revision 11-Aug-1995 #text_change 22-Jun-1999

C;Accession: A56171; S02809; S71395

R;Catasus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.

J. Biol. Chem. 270, 6651-6657, 1995

A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA

A;Reference number: A56171; MUID:95204457; PMID:7896805

A;Accession: A56171

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-417 <CAT>

A;Cross-references: GB:U019977; NID:G790226; PIDN:AAV4425.1; PID:G790227

A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340

R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.

Eur. J. Biochem. 179, 609-616, 1989

A;Title: Purification and properties of five different forms of human procarboxypeptidase

A;Reference number: S02809; MUID:89153096; PMID:2920728

A;Accession: S02809

A;Molecule type: protein

A;Residues: 'S',18-36,'N',38-43 <PAS>

R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walcott

Arch. Biochem. Biophys. 332, 8-18, 1996

A;Title: Expression and characterization of human pancreatic procarboxypeptidase A1

A;Reference number: S71394; MUID:96400327; PMID:8806703

A;Accession: S71395

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-301,'T',303-338,'SR',341-417 <LAE>

C;Genetics:

A;Gene: GDB:CPA2

A;Cross-references: GDB:125230; OMIM:600688

A;Map position: 7q32-qter

C;Superfamily: carboxypeptidase

C;Keywords: hydrolase; metallo-carboxypeptidase; zinc

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-110/Domain: activation peptide #status predicted <ACP>

F;Fill-417/Product: carboxypeptidase A2 #status predicted <MAT>

F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted

F;246-269/Dsulfide bonds: #status predicted

F;356,378/Active site: Tyr, Glu #status predicted

Query Match 26.6%; Score 508; DB 2; Length 417;

Best Local Similarity 32.5%; Pred. No. 5.6e-34;

Matches 128; Conservative 63; Mismatches 137; Indels 66; Gaps 11;

QY 19 HVF---AFQSGQVLAALPRTSRQVQLQNLTTTVEIVL--WQPVTAFLVKKQVHFFVN 73

DB 13 HIYCLETFVGQVLEIVPSNEEQIKNLLQLEAQEHLQDLDFWKSPT---TPGETAHVRVP 68

QY 74 ASDVNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS---YYQYHSLNEIY 130

DB 69 FVNQAVKVFLESQGIAYSIMIEDVQLLDKENEEMLEFRNRERRSGNFNGAYHTLEIS 128

QY 131 SWIEFITERHDPDLTKIHGSSPEKYPYLVKVS--GKQETAKNAIWDGTHAREWISPA 189

DB 129 QEMDNLVAEHPGLVSKVNIQSGFENRPMNVILKFTSGGD---KPAIWLDAIGHAREWVQA 185

QY 190 FCLWFIHG-----HNRWKNRSFYANNH 212

DB 186 TALWTANKIVSDYGDPSITSLDALDIFLPTVNTDGVFVSQTKRMRKTRSKVSGSL 245

QY 213 CIGTDLNSFVSKHWCCEAGSSSCSETYCGLYPESEPEVKAFLRNINQIKAYISM 272

DB 246 CVGVDPNPNW-DAGFGPGASSNPNPCSDYHGSPANSSEVEVKSIVDFIKSH-GKVAFIIL 303

QY 273 HSYQSHIVFPYSTRSKSDHEELSLVASEAVRAIDTKSKTRYTHGHSSTLYLAPGG 332

DB 304 HSYQLLMFPYPYKCTKLDLDFELSEVQAQASL-RSLHGHTKYKVGKPGICSVIYQASGS 362

QY 333 DDWIVDLGIKYSFTSN-----PPVEKLLP 356

DB 363 IDWSYDYGKYSFAFELRDGTGRYGLLPARQILP 396

RESULT 12

S29127

carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human

N;Alternate names: pancreatic carboxypeptidase A1

C;Species: Homo sapiens (man)

C;Date: 25-Feb-1994 #sequence revision 19-Jan-1996 #text_change 18-Jun-1999

C;Accession: S29127; A34205; S02810; S71394; S02811

R;Catasus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver

Biochem. J. 287, 299-303, 1992

A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.

A;Reference number: S29127; MUID:93038569; PMID:1417781

A;Accession: S29127

A;Molecule type: mRNA

A;Residues: 1-419 <CAT>

A;Cross-references: EMBL:X67318; NID:G35329; PIDN:CAA47732.1; PID:G35330

R;Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.

Am. J. Hum. Genet. 46, 795-800, 1990

A;Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.

A;Reference number: A34205; MUID:90196012; PMID:1969228

A;Accession: A34205

A;Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA
A:Residues: 330-396 <STE>
A>Note: the authors translated the codon CTG for residue 391 as Val
R:Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
PDBS Lett. 261, 179-183, 1990
A>Title: Further studies on the human pancreatic binary complexes involving procarboxypeptidase A
A:Reference number: S08253; MUID:90169111; PMID:2307232
A:Accession: S08253
A:Molecule type: protein
A:Residues: 17-43; 'XX', 114-135 <MOV>
R:Paacual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A>Title: Purification and properties of five different forms of human procarboxypeptidase A
A:Reference number: S02809; MUID:89153096; PMID:2920728
A:Accession: S02810
A:Molecule type: protein
A:Residues: 17-42 <PAS>
R:Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton Arch. Biochem. Biophys. 332, 8-18, 1996
A>Title: Expression and characterization of human pancreatic procarboxypeptidase A1
A:Reference number: S71394; MUID:96400327; PMID:8806703
A:Accession: S71394
A:Status: not compared with conceptual translation
A:Residues: 1-419 <LAB>
A:Molecule type: mRNA
A:Genetics: 1-419 <LAB>
A:Gene: GDB:CPA1; CPA
A:Cross-references: GDB:120597; OMIM:114850
A:Map position: 7q32-7qter
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status predicted <ACP>
F:111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F:179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F:248-271/Disulfide bonds: #status predicted
F:358,380/Active site: Tyr, Glu #status predicted

Query Match 25.6%; Score 489.5; DB 1; Length 419;
Best Local Similarity 32.7%; Pred. No. 1.9e-32;
Matches 131; Conservative 60; Mismatches 127; Indels 83; Gaps 13;

Qy 3 LCSLAVLPVILFCQHVFAPQSGOVLAALPRTSQOVQLQNTTYEIVL--WQ----- 55
Db 4 LVLVSLVGVGVGKED----FVGHQVLRISVADEAQVQKVELEDLEHLQDFWRGPAHP 59
Qy 56 --PVTADLVKKQVHFFPNASVDVNVKAHLNVSGIPCSVLADVEDLI---QQQI----- 106
Db 60 GSPTDV-----RVPP-----PSIQAVKIFLESHGISVETWIEDVQSLDDEEQQMFAPR 108
Qy 107 ----SNDTVSPRASAYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLK 162
Db 109 SRARSTDTFN-----YATYHLEEIYDFDLVAENPHLVSKIQIGNTYEGRPVLYLK 161
Qy 163 VSGKEQAKNAIWDGCIHAREWISPAFLWFI----- 195
Db 162 FS-TGGSKRPAIWDITGHSREWVTQASGVWFAKKITQYQDAFTAAILDTLDFLEIV 220
Qy 196 -----GHNRMWRNRSFYANNHCGITDLNSFVKHWCCEGASSSCSEYCYGLY 245
Db 221 TNPQGFATHSTNRWRTKTRHTAGSLCIGVDPNRNW-DAGFGLSGASSNPCSEYTHGKF 279
Qy 246 PESEPEVKAVASFLRRNINQIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVR 305
Db 280 ANSEVEKSIIVDFVKDHCN-IKAFISIHYSQQLMYPYGYKTEPVDPDQDELQLSKAATV 338
Qy 306 AIDTKSKNTRYTHGSGSTLYLAPGGDDWYDLGKIYSFT 346
Db 339 AL-ASLYGCKENYGSIIKAIYQASGSTIDWTYSQGIKYSFT 378

RESULT 13
T33527

hypothetical protein T06A4.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C:Accession: T33527
R:Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid T06A4.
A:Reference number: Z21384
A:Accession: T33527
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-528 <WUX>
A:Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A:Experimental source: strain Bristol N2; clone T06A4.
C:Genetics:
A:Gene: CESP:T06A4.1
A:Map position: 1
A:Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C:Superfamily: carboxypeptidase

Query Match 22.9%; Score 437.5; DB 2; Length 528;
Best Local Similarity 32.3%; Pred. No. 4.8e-28;
Matches 120; Conservative 57; Mismatches 122; Indels 72; Gaps 14;

Qy 33 PRTSRQVQLNL---TTTYEIVLWQPVY-----ADLVKKKQVHFFPNASVDVNVKAHL 84
Db 54 PETEGSVKYLRSYLEDPSYELDFMQPPTNIGAIVDLTAVAPADAPFV--KDLSEKK--- 108
Qy 85 NVSGIPCSVLADVEDLIQQQISNDTVSPRASAYEQYHSLNEIYSWIEFITERHPDML 144
Db 109 ----ISYIVAVNDLSKALENERGSKDFYNVAGFAIDYKINSLDEEIQTEMKRLKKSYPTMI 164
Qy 145 TKIHGSSFEKYPVLYLVKSGKEQT--AKNAIWDGCIHAREWISPAFLWFI----- 197
Db 165 TLIDIGQSHENRTLLVMKITGRNPLGSKISNWDAGIHAREWIATATAM-YIAHELLLG 223
Qy 198 -----NRMWRNRS-----FYANNHICIGTD 217
Db 224 YENDATVAKLMDHIDFYLPMNPDGYEYSREKNRMWRNRSAPKARQTFSTVCCSGVD 283
Qy 218 LNSFVSKHWCCEGASSSCSEYCYGLYPEPEVKAVASFLRRNINQIKAYISMHSYSQ 277
Db 284 LNRNP-DWFMASTGSSSDPCHDTYHGSAAFPSEPSQAVRDLFQNTPE--AFISLHYSQ 340
Qy 278 HIVFPYSYTR-SKSKD-HEELSLVASEAVRAIDTKSKNTRYTHGSGSTLYLAPGGDDW 335
Db 341 MWLPYGRKQSYQDYHTGLRPLALRATKALYEL-YGTYQVGTGADLMYEASGSHDW 399
Qy 336 IY-DLGKIYSF 345
Db 400 AKGQLKVPYAY 410

RESULT 14
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T20507; T24289
R:Matthews, L.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19284
A:Accession: T20507
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-323 <WIL>
A:Cross-references: EMBL:Z78411; PIDN:CAB01647.1; GSPDB:GN00023; CESP:F02D8.4
A:Experimental source: clone F02D8
R:Wild, A.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19869
A:Accession: T24289
A:Status: preliminary; translated from GB/EMBL/DBDJ

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:03:36 ; Search time 14.2797 Seconds
(without alignment)
1185.568 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPVLFCEQHV.....IKYSFTNPVPVEKLLPLSLK 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	692	36.2	415	1	CBPB RAT
2	641	33.5	417	1	CBPB HUMAN
3	628.5	32.9	401	1	CBPB DIG
4	628.5	32.9	401	1	CBPB MOUSE
5	627	32.8	416	1	CBPB CANFA
6	619.5	32.4	417	1	CBPC HUMAN
7	580	30.4	306	1	CBPB BOVIN
8	553	28.9	417	1	CBP2 RAT
9	546	28.6	309	1	CBPC RAT
10	533.5	27.9	419	1	CBPA BOVIN
11	529.5	27.7	419	1	CBP1 RAT
12	511	26.7	417	1	CBP2 HUMAN
13	489.5	25.6	419	1	CBP1 HUMAN
14	483	25.3	421	1	CBP4 HUMAN
15	404.5	21.2	304	1	CBP2 SIMVI
16	386	20.2	433	1	CBPA ANOGA
17	369.5	19.3	303	1	CBPB ASTFL
18	299.5	15.7	430	1	XHT2 YEAST
19	277.5	14.5	424	1	CBPT THEUV
20	271	14.2	451	1	CBPS STRGP
21	251	13.1	434	1	CBPS STRCP
22	118.5	6.2	707	1	ORC1 SCHPO
23	114.5	6.0	1380	1	CBPD HUMAN
24	112.5	5.9	1377	1	CBPD MOUSE
25	105	5.5	1255	1	PER2 HUMAN
26	103.5	5.4	722	1	CPXM MOUSE
27	102	5.3	734	1	CPXM HUMAN
28	100	5.2	376	1	YOGT BACSU
29	99.5	5.2	445	1	TPH1 CHICK
30	98.5	5.2	488	1	TY3H ANGAN
31	97	5.1	663	1	TAZ1 SCHPO
32	95	5.0	444	1	TPH1 HUMAN
33	91.5	4.8	465	1	STHA_ECO57

34	91.5	4.8	465	1	STHA_ECOLI
35	91.5	4.8	3588	1	SRF1_BACSU
36	91	4.8	297	1	Y103 HUMAN
37	91	4.8	444	1	TPH1_RABIT
38	91	4.8	982	1	ENV_SFV3L
39	90	4.7	1451	1	SPT6 YEAST
40	89.5	4.7	682	1	CJBA_BACUH
41	89	4.7	525	1	PUR9_BUCAI
42	88.5	4.6	319	1	UN30 CAEL
43	88.5	4.6	466	1	STHA_SALTY
44	88	4.6	454	1	CBPH_LOPAM
45	88	4.6	1150	1	C9EA_BACTA

ALIGNMENTS

RESULT 1
CBPB RAT
ID CBPB RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
XP [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
genes. Comparative analysis of the rat carboxypeptidase gene
family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine (or L-arginine) + H(2)O =
peptide + L-lysine(or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; M23959; AAA40872.1; JOINED.
EMBL; M23947; AAA40872.1; JOINED.
EMBL; M23950; AAA40872.1; JOINED.
EMBL; M23952; AAA40872.1; JOINED.
EMBL; M23953; AAA40872.1; JOINED.
EMBL; M23954; AAA40872.1; JOINED.
PIR; A32129; A32129.
HSSP; P09955; 1NSA.
MEROPS; M14.003.
InterPro; IPR003146; Propep M14.
InterPro; IPR000834; Zn carboxepte.
Pfam; PF02244; Propep M14; 1.
Pfam; PF00246; Zn carboxepte; 1.
SMART; SM00631; Zn pept; 1.
PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 13
FT PROPEP 14 108 ACTIVATION PEPTIDE.
FT CHAIN 109 415 CARBOXYPEPTIDASE B.
FT METAL 174 174 ZINC (BY SIMILARITY).
FT METAL 177 177 ZINC (BY SIMILARITY).
FT METAL 302 302 ZINC (BY SIMILARITY).

FT ACT SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 171 184 BY SIMILARITY.
FT DISULFID 243 266 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
SQ SEQUENCE 415 AA; 47515 MW; 8EA06CCADE30B6F2 CRC64;
Query Match 36.2%; Score 692; DB 1; Length 415;
Best Local Similarity 39.2%; Pred No. 4.1e-51;
Matches 150; Conservative 61; Mismatches 126; Indels 46; Gaps 7;
QY 1 MKLCSLAVLVPILVPCQHVFAFQSGQVLAALPRTSRQVQLNLTYYEIVLMQPVAD 60
Db 1 MLLLLALVSVLAHASEEH---FDGNRVYRVSVHGEDHVNLIQELANTKEIDFWKPDGAT 57
QY 61 LIVKKQVHFFVNASVDNVAHLNVSGIPCSVLLADVEDLIQQQISNDTSPRASASY 120
Db 58 QVKPLTVDVDFHKAEDVADVENFLENEVHYELISNVRNALESQDSHT---RASGHSY 114
QY 121 EYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPVLYLVKVSKEQTAKNAIWDGCI 180
Db 115 TYKNWETIEAWIQVATDNDPLVQSVIGTFEGRNMYVLKI-GKTRPNKPAIFIDCGF 173
QY 181 HAREWISPAFLWFI-----GH-----NRMRK 203
Db 174 HAREWISPAFCQFVREAVRTYNOBIHMKQLLDLDFVLPVWNIDGYVYTWTKDRMRK 233
QY 204 NRSFYANNHCITDLSNFVSKHWCBEAGSSCSETVCGLYPESEPEVKAVASFLRNI 263
Db 234 TRSTWAGSSCLGVRENRF-NAGWCEVGASRSPCSETYCGPAPESEKETKALADPIRNL 292
QY 264 NOIKAYISNHSYQHVPPYSYTRSKSDHELSLVASEAVRAIDKTSNTRYTHGHS 323
Db 293 STIKAYLTHSQQMQLFYSDYKLPENYEELNALVGAKEEL-ATLHGKITYGPGAT 351
QY 324 TLYLAPGGDDWYDLGKYSPT 346
Db 352 TIYPAAGSDWSDYQGIKYSPT 374
RESULT 2
ID CBPB HUMAN STANDARD; PRT; 417 AA.
AC P15086; O60834; Q96B08;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase B precursor (BC 3.4.17.2) (Pancreas-specific protein) (PASP).
DE CPB1 OR CPB OR PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC TISSUE=Pancreas;
RX MEDLINE=92129345; PubMed=1370825;
RA Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S., French C.K.;
RT "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B.";
RT J. Biol. Chem. 267:2575-2581(1992).
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 110 CARBOXYPEPTIDASE B.
FT CHAIN 111 417 ZINC (BY SIMILARITY).
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 259 268 BY SIMILARITY.
FT DISULFID 259 273 BY SIMILARITY.
FT CONFLICT 16 16 H -> A (IN REF. 1; AA SEQUENCE).
FT CONFLICT 17 17 H -> Q (IN REF. 1; AA SEQUENCE).

RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K., Altschul S.F., Collins B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S.J., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 16-43.
RC TISSUE=Pancreas;
RX MEDLINE=89153096; PubMed=2920728;
RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT "Purification and properties of five different forms of human procarboxypeptidases.";
RT Eur. J. Biochem. 179:609-616(1989).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-lysine (or L-arginine) + H(2)O = peptide + L-lysine (or L-arginine).
CC -!- TISSUE SPECIFICITY: Pancreas.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81057; AAA66973.1; -.
CC EMBL; AJ224866; CAAL2163.1; -.
CC EMBL; BC015338; AAH15338.1; -.
CC PDB; 1KWM; 28-AUG-02.
CC MEROPS; M14.003; -.
CC Genew; HGNC:2299; CPB1.
CC MIM; 114852; -.
CC GO; GO:0004180; F:carboxypeptidase activity; TAS.
CC InterPro; IPR001146; Propep_M14.
CC InterPro; IPR000834; Zn_carbopept.
CC Pfam; PF02244; Propep_M14; 1.
CC Pfam; PF00246; Zn_carbopept; 1.
CC SMART; SM00631; Zn_pept; 1.
CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
FT 3D-structure.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 110 CARBOXYPEPTIDASE B.
FT CHAIN 111 417 ZINC (BY SIMILARITY).
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 259 268 BY SIMILARITY.
FT DISULFID 259 273 BY SIMILARITY.
FT CONFLICT 16 16 H -> A (IN REF. 1; AA SEQUENCE).
FT CONFLICT 17 17 H -> Q (IN REF. 1; AA SEQUENCE).


```
FT CONFLICT 37 37 H -> Q (IN REF. 4).
FT CONFLICT 208 208 D -> N (IN REF. 2).
FT CONFLICT 245 245 MISSING (IN REF. 1).
SQ SEQUENCE 417 AA; 47367 MW; EBBB98B27F5D5AF9 CRC64;

Query Match 33.5%; Score 641; DB 1; Length 417;
Best Local Similarity 35.3%; Pred No. 8.6e-47;
Matches 135; Conservative 71; Mismatches 132; Indels 44; Gaps 7;

QY 3 LCSLAVLPVILFCBQHVA-FQSQVLAALPRTSRQVQLNQLTTTYEIVLMQPVTLADL 61
Db 1 MLALLVLTVALASAHGGEHGEKVPFVNVEDENHINIIRELASTTQIDFWKPDSTVQ 60
QY 62 IVKKQVFFVNASVDNVKHLNVGIPCSVILLADVEDLIQQQISNDTVPASASYE 121
Db 61 IKPSTVDVFRKAEDTIVTVENLVKQELQYKVLISNLRNVVEAQFDSRV---RATGHSYE 117
QY 122 QVHSLNELYSMTETTERHPDMLTKIHGSSPEKYPVLYLVKSGKEQAKNAIWDCCGIH 181
Db 118 KINKWETTEATQOVATENPALISRSVIGTTFEGRAIYLLKV-KGAGQNKPAIFMDCGFH 176
QY 182 AREWISPAFLWFI-----GH-----NRMWRKN 204
Db 177 AREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLVPLNIDGIYITWTKSRFMT 236
QY 205 RSFYANNHICIGTDLNSNFVSKWCEGASSSCSETYCGLYPESPEVKAVASFLRRNIN 264
Db 237 RSTHTGSSCIGTDPNRF-DAGWCIGASRNPDETYCGPAESEKETKALADFTIRNKL 295
QY 265 QIKAYISHSYOHIVFPVYSTRSKSDHEELSLVASEAVRAIDKTSKNTVTHGSGSET 324
Db 296 SIKAYLTHSYQMMIYIPSYAYKUGENNAELNALAKATVKEL-ASLHGTYKTYGPGATT 354
QY 325 LYLAPGGDDMIYDGIKYSFT 346
Db 355 IYPAGGSDWAYDQGIYSFT 376

RESULT 3
ID CBPB_PIG STANDARD; PRT; 401 AA.
AC P09955;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (BC 3.4.17.2).
GN CPB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
RN [1]
RP SEQUENCE OF 1-103.
RX MEDLINE=91208150; PubMed=2018774;
RA Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA Aviles F.X.;
RT "Analysis of the activation process of porcine procarboxypeptidase B
RT and determination of the sequence of its activation segment.";
RL Biochemistry 30:4082-4089(1991).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-38.
RX MEDLINE=85279427; PubMed=4026847;
RA Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT "Sequential homologies between procarboxypeptidases A and B from
RT porcine pancreas.";
RL Biochem. Biophys. Res. Commun. 130:97-103(1985).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=91114690; PubMed=1989878;
RA Coll M., Guasch A., Aviles F.X., Huber R.;
RT "Three-dimensional structure of porcine procarboxypeptidase B: a
RT structural basis of its inactivity.";
RL EMBO J. 10:1-9(1991).
```

```
RN [4]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX MEDLINE=91027767; PubMed=2223783;
RA Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RT "Sequence-specific 1H NMR assignments and determination of the
RT secondary structure for the activation domain isolated from
RT pancreatic procarboxypeptidase B.";
RL Biochemistry 29:7515-7522(1990).
RN [5]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX MEDLINE=91114693; PubMed=1989879;
RA Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
RT "The NMR structure of the activation domain isolated from porcine
RT procarboxypeptidase B.";
RL EMBO J. 10:11-15(1991).
RN [6]
RP STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX MEDLINE=93044373; PubMed=1422143;
RA Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RA Huber R., Wuethrich K.;
RT "Comparison of the NMR solution structure with the X-ray crystal
RT structure of the activation domain from procarboxypeptidase B.";
RL J. Biomol. NMR 2:1-10(1992).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-lysine (or L-arginine) + H(2)O =
CC peptide + L-lysine (or L-arginine).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -I- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR PDB; 1PBA; 31-OCT-93.
DR PDB; JNSA; 24-DEC-97.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW 3D-structure.
FT PROPEP 1 95 ACTIVATION PEPTIDE.
FT CHAIN 96 401 CARBOXYPEPTIDASE B.
FT DISULFID 158 171
FT DISULFID 230 253
FT DISULFID 244 258
FT METAL 161 161 ZINC.
FT METAL 164 164 ZINC.
FT METAL 289 289 ZINC.
FT ACT_SITE 341 341
FT ACT_SITE 363 363 NUCLEOPHILE.
FT TURN 8 9
FT STRAND 11 17
FT HELIX 20 32
FT STRAND 36 39
FT HELIX 43 45
FT STRAND 50 56
FT HELIX 58 60
FT HELIX 61 69
FT TURN 70 72
FT STRAND 74 79
FT HELIX 82 87
FT TURN 88 88
FT TURN 101 102
FT HELIX 107 120
FT TURN 122 124
FT STRAND 125 132
FT TURN 134 135
FT STRAND 138 144
FT TURN 149 150
FT STRAND 153 157
FT TURN 162 163
FT HELIX 165 181
```



```

TURN 182 184
FT HELIX 186 194
FT STRAND 196 200
FT STRAND 205 213
FT TURN 214 217
FT TURN 216 217
FT STRAND 224 224
FT TURN 226 227
FT STRAND 231 231
FT HELIX 235 237
FT TURN 243 244
FT STRAND 249 249
FT TURN 252 253
FT TURN 255 256
FT STRAND 257 257
FT TURN 262 263
FT HELIX 266 277
FT TURN 278 281
FT STRAND 282 289
FT STRAND 294 297
FT TURN 307 308
FT HELIX 309 327
FT STRAND 332 335
FT HELIX 336 339
FT HELIX 347 353
FT TURN 354 355
FT STRAND 359 363
FT TURN 371 372
FT HELIX 376 378
FT HELIX 379 398
FT TURN 399 401
SQ SEQUENCE 401 AA; 45713 MW; 53129AF159A26348 CRC64;

Query Match 32.9%; Score 629.5; DB 1; Length 401;
Best Local Similarity 37.1%; Pred. No. 7.7e-46;
Matches 134; Conservative 64; Mismatches 120; Indels 43; Gaps 6;

QY 23 FQSGOVLAALPRTSRQVQLNLTYYEIVLVQPTADLIYKVKQVHFVNASVDVNYKA 82
DB 7 FEQGVFVNVNEDNDISLHELASTRQIDFWKPSVTQIKPHSTVDVFRVKAEDILAVED 66

QY 83 HLNVSIGPCSVLLADVEDLIQQISNDTVSPRASAYVEQVHSLNEIYSWIFETERHPD 142
DB 67 FLEQNELQVLEVLNLSRSLVLEAQFDSRC---RTTGHSEYKYNWETIEAWTQVTSKPD 123

QY 143 MLTKTHIGSFPKYPFLYLVKVGSKQTAKNAIWDICGTHAREWISPAFLWFI----- 195
DB 124 LLSRGAIGTTFDGDNIYLLKV-GKPGSNKPAIFMDCGFHAREWISQAFQCFVROAVRTY 182

QY 196 -----GH-----NRWRKNRSFYANNHCIGTDLNSNFYSK 225
DB 183 GYEAHMTFELNDLFYVLPVLNIDGIYTWTKNRMWKRTRSTNAGSCCTGTDPNRNF-NA 241

QY 226 HWCEBGASSSCSYCYGLYPESEPEVKAVASFLRRNINQIKAYISMSHSYOHIVFPYSY 285
DB 242 GCHTVGASVNPCEYCYGSAASEKETALADFIRNLSIKAYITIHYSQWILYPSY 301

QY 286 TRSKDHLESLVASEAVRAIDKTSKNTRYTHGSETFLYLAPOGGDDWIYDLGKYSF 345
DB 302 DYKLPENDAEFLSLAKGAVKEL-ASLYGTSTSYGPGSTIYPAAGSDDDWAYNQIKYSF 360

QY 346 T 346
DB 361 T 361
```

RESULT 4

```

ID CBPC_MOUSE STANDARD; PRT; 417 AA.
AC P15089;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Serafin W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RL novel member of the carboxypeptidase gene family.";
RJ J. Biol. Chem. 264:20094-20099(1989).
CC -|- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -|- SUBCELLULAR LOCATION: Secretory granules.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DE EMBL; J05118; AAA37369.1; -.
DR PIR; A34487; A34487.
DR HSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR MGI; 88479; Cp3.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF02246; Zn_carbopept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
SQ SEQUENCE 417 AA; 48790 MW; A2B300A068D1BA6D CRC64;

Query Match 32.9%; Score 628.5; DB 1; Length 417;
Best Local Similarity 37.0%; Pred. No. 9.8e-46;
Matches 140; Conservative 60; Mismatches 131; Indels 47; Gaps 8;

QY 6 LAVLVPIVLFCEQHFQAFQSQGVLAALPRTSRQVQLNLTYYEIVLVQPTADLIYKVK 65
DB 7 MAVIYTTTLAIAPVH---FDREKVPFRVKLQNEKHSVLKSLTQSIELDFWYDAIDVAVN 63

QY 66 KQVHFVNASVDVNVKAHLNVSGIPCSVLLADVEDLIQQIS-NDTVSPRASASYVEOVH 124
DB 64 MTVDPRVSEKESQTIQSTLEQKHIEYLIDHLEQEEIKQDFDKDEIAGRIS---YAKYN 120

QY 125 SLNEIYSWIFETERHPDMLTKIHIGSFPKYPFLYLVKVGSKQTAKNAIWDICGTHARE 184
DB 121 DWDKIVSWTEKLEKHPWVSRIKIGSTVEDNPLVVLKI-GKDKGERKAIWMDCGIHARE 179

QY 185 WISPAFLWFI-----GHN-----RWRKNRSF 207
DB 180 WISPAFCQWFFYQATKYSYKGNKIMTKLLDRMNFYVLPVFNVDGIYTSWTQDRMRKNRSR 239

QY 208 YANNHCIGTDLNSNFVSKHWCEGASSSCSYCYGLYPESEPEVKAVASFLRRNINQIK 267
```

```

Db 240 NONSTCIGDNLNRF-DVSWDSSPTNKPCLNVRGPAPSEKVKAVTNFIRSHLSIK 298
QY 268 AYISHMSYQHIVFPYTSRKSQHEELSLVASEAVRAIDKTSKNTRYTHGHSGSETLYL 327
Db 299 AYITHSYSQMLLIPYGYTFKLPPNHQDLKLVARIATDAL-STRYETRYIYGPIASTIYK 357
QY 328 APGGDDWIYDLGIKYSF 345
Db 358 TSGSLDWDYDLGIKHTF 375

RESULT 5
CBPC CANFA STANDARD; PRT; 416 AA.
AC P55261;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE membrane associated protein) (ZAP47).
GN CPB1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Fukuoka S.-I.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =
CC peptide + L-lysine (or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D78348; BA011366.1; -.
DR HSSP; P09955; 1PBA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboPept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboPept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 110 416 CARBOXYPEPTIDASE B.
FT METAL 175 175 ZINC (BY SIMILARITY).
FT METAL 178 178 ZINC (BY SIMILARITY).
FT METAL 303 303 ZINC (BY SIMILARITY).
FT ACT SITE 377 377 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 244 267 BY SIMILARITY.
FT DISULFID 258 272 BY SIMILARITY.
SQ SEQUENCE 416 AA; 47595 MW; DFEL1992CDS2F8AB4 CRC64;

Query Match 32.8%; Score 627; DB 1; Length 416;
Best Local Similarity 35.5%; Pred. No. 1.3e-45;
Matches 135; Conservative 68; Mismatches 127; Indels 50; Gaps 8;

QY 8 VLVPIVL-----PCEQHVFAFGSQGVLAALPRTSRQVQLVQLNTTYTVILVQPVTDILIV 63
Db 5 ILVTLALASAHYSGEH---FEQEGFRVNVEDENHINLLTLASTQIDFWKPKDSVTQIK 61

```

```

QY 64 KKKQVHFVFNASDVNDVNVKAHLNVSGIPCSVLADVEDLIQQQISNDTVSPRASASYEQY 123
Db 62 PHSTADPRVKAEDILTVDLQKQNELHYEVLNNLRLLVLEGQFGRQV---PATGHSYKY 118
QY 124 HSLNEIYSWIFETTERHEDMLTKIHGSSFEKYPYLVKVSKEQTAKNAIWDGIIHAR 183
Db 119 NRWETIEATQCVTSNPDLISRRSIGTTFEGRTIYLLKV-GKAGQNKPAIFMDCGPHAR 177
QY 184 EWISPAFLCFEI-----GH-----NRWKNRS 206
Db 178 EWISPAFWQVFXRIRTYQGEIHWTLLDKLDFVLPVGNIDGVYVYTWKRMKRTS 237
QY 207 FYANNHCIGTDLNPNFVSKHWCEGASSSSCSETYCGLYPSEPEPVKAVASFLLRNINOI 266
Db 238 TQVGTNCVGTDPTRNF-DAGWCKIGASRNPCEDETCGPAESEKETKALANFIRNLSI 296
QY 267 KAYTSMHSYSQHIYVPPISYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHSETLY 326
Db 287 KAYLTIHSYQWMLYPYSYDYKLTENNAELNALAKATVKEL-ATLHGTYKYTPGCGATTY 355
QY 327 LAPGGDDWIYDLGIKYSFT 346
Db 356 PAAGSSDDWAYDQGIKYSFT 375

RESULT 6
CBPC HUMAN STANDARD; PRT; 417 AA.
AC P15088; Q96E94;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=90083291; PubMed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugartaker D.J.,
RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mast cells;
RX MEDLINE=92105393; PubMed=179276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A."
RL J. Clin. Invest. 89:273-282(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buéto K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```



```

RL J. Biol. Chem. 244:5246-5253(1969).
RP ACTIVE SITE.
RX MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RT "Identification of a glutamic acid at the active center of bovine
RT carboxypeptidase B."
RL J. Biol. Chem. 247:7864-7869(1972).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR: A93797; CPBOB.
DR PDB: 1CPB; 30-SEP-83.
DR MEROPS: M14.003; -.
DR InterPro: IPR000834; Zn_carboPept.
DR Pfam: PF00246; Zn_carboPept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT METAL 149 163
FT METAL 66 66
FT METAL 69 69
FT METAL 194 194
FT ACT SITE 246 246
FT ACT SITE 268 268
SQ SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;

Query Match 30.4%; Score 580; DB 1; Length 306;
Best Local Similarity 43.6%; Pred. No. 8.2e-42;
Matches 115; Conservative 44; Mismatches 65; Indels 40; Gaps 5;

QY 120 YEYHSLNEIYGIWIFETRHDPDLTKIHGSSFEKYPVLYLVKSGKEOTAKNAIWDIG 179
DB 6 YEKYNWETIETAWTQVASENPDLISRAITFLGNITLYLVK-GKPSNKPAPVMDCG 64

QY 180 IHAREWISPAFLWFI-----GH-----NRMWR 202
DB 65 PHAREWISPAFCQWFVREAVRTYGREIHWTEFLDKLDFVLPVFNIDGVIYTWTTNRMWR 124

QY 203 KRSFYANNHCGTDLNLFVSKHCEGASSSCSETYCGLYPESEPEVKAVASFLRN 262
DB 125 KTRSTRAGSCCTGDLNRF-DAGWCISGASNPCCSETYCGSAESESKEKAVADFINH 183

QY 263 INQIKAYISMHSYOHVPPYSTRSKDHELSIVASEAVRAIDKTSKNTRYTHGHS 322
DB 184 LSSIKAYLIHYSQMLPISYIDYKLPKNNVELNLAKGAVKKL-ASLHGTTYSYGPA 242

QY 323 EFLYAPGGDDWIYDLGIKYSFT 346
DB 243 TTIIYASGGSDMDAYDQGIKYSFT 266

RESULT 8
CBP2_RAT ID_CBP2_RAT STANDARD; PRT; 417 AA.
AC P19222;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=89034323; PubMed=3182871;
RA Gardell S.J., Craik C.S., Clause E., Goldsmith E.J., Stewart C.B.,
```

Db 58 ----IPGETVHVVPFASIQAVKVFLESQIDYSIMIEDVQLLDOEREEMLFNOQERNG 113
QY 116 SASVYEOVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPLYVLKYS-GKEQAKNAI 174
Db 114 GNFNFEAYHTLEIYQEDMNLVAENPGLVSKVNLGSSPENNPMNLKFTSGD--KPAI 170
QY 175 WDCGHHAREWISPAFLM-----FI-----GH----- 197
Db 171 WLDAGIHAREWTOATALTANKIASDYGTDPAITSLNTLIDIFLLPVTNPDGYVPSQTT 230
QY 198 NRMWRKRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESEPEVKAVAS 257
Db 231 NRMWRKRSKSGSGCVGVDPNRNW-DANFGGPGASSPCSDSYHGPKPENSEVEKSIDV 289
QY 258 FURRNINQKAVISMSHSYQHVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYT 317
Db 290 FIKSH-GKVKAPITLHYSQQLMFPYGYKTRPDDFDELVAQAAQAL-KRLHGTYSK 347
QY 318 HGHGSETLVAPGGDDMTYDLGIKYSFTSN-----PPVEKILP 356
Db 348 VGPICSVIYQASGSDIWDYDLGIKYSFAFELRDTAFYGFLLPAKOILP 396

RESULT 9
CBPC RAT
ID CBPC RAT STANDARD; PRT; 309 AA.
AC P21961;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase, A3).
DE CPAS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA Neurath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules.";
RT Biochemistry 30:648-655(1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A38395; A38395.
DR HSSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR InterPro; IPR000834; Zn carbopept.
DR Pfam; PF0246; Zn carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydroxylase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 196 196 ZINC (BY SIMILARITY).
FT ACT_SITE 248 248 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 270 270 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 65 78 BY SIMILARITY.
FT DISULFID 137 160 BY SIMILARITY.
FT VARIANT 1 1 MISSING (IN THE MAJOR FORM).
SQ SEQUENCE 309 AA; 35786 MW; 20330FABC3EE83EF CRC64;
Query Match 28.6%; Score 546; DB 1; Length 309;
Best Local Similarity 42.2%; Pred. NO. 6.3e-39;
Matches 111; Conservative 39; Mismatches 73; Indels 40; Gaps 5;

QY 120 YEOVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPLYVLKYSKQBOTAKNAIWDCG 179
Db 8 YAKYNDWKNKIVSWTEKWKVEKPEWVRKIKIGSTVEDNPLYVLKI-GRKDGKKAIPMDCG 66
QY 180 IHAREWISPAFLMFI-----GH----- 202
Db 67 IHAREWISPAFLMFI-----GH----- 126
QY 203 KNRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 262
Db 127 KNRSKNPSTCIGTDLNENF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIKSH 185
QY 263 INQIKAYISMSYSQHVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHS 322
Db 186 LNSIKAYITFHSYQQLMFPYGYKTRPDDFDELVAQAAQAL-KRLHGTYSK 244
QY 323 ETLVLAPGGDDMTYDLGIKYSF 345
Db 245 STIYKTSGLSDWAYDGLGKHTF 267
RESULT 10
CBPA BOVIN
ID CBPA BOVIN STANDARD; PRT; 419 AA.
AC P00730;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase A precursor (EC 3.4.17.1).
DE CPA.
GN CPAS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91151335; PubMed=198496;
RA le Huerou I., Guilloteau F., Toullec R., Puigserver A., Wicker C.;
RT "Cloning and nucleotide sequence of a bovine pancreatic
RT preprocarboxypeptidase A cDNA.";
RL Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RX MEDLINE=96096549; PubMed=8522204;
RA Goo J.H., Kim K.H., Choi K.Y.;
RT "Cloning, sequencing and expression of the gene encoding a major
RT allotypic preprocarboxypeptidase A from bovine pancreas.";
RL Gene 165:333-334(1995).
RN [3]
RP SEQUENCE OF 111-417.
RX MEDLINE=71155183; PubMed=5102489;
RA Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RT chymotryptic peptides of the cyanogen bromide fragment F-1.";
RL Biochemistry 10:938-950(1971).
RN [4]
RP REVISIONS TO 138 AND 141.
RX MEDLINE=72138789; PubMed=5143102;
RA Petra P.H., Hermanson M.A., Walsh K.A., Neurath H.;
RT "Characterization of bovine carboxypeptidase A (Allan).";
RN [5]
RP SEQUENCE OF 17-120.
RX MEDLINE=89150306; PubMed=3147705;
RA Wade R.D., Hase G.M., Kumar S., Walsh K.A., Neurath H.;
RT "The amino acid sequence of the activation peptide of bovine pro-
RT carboxypeptidase A.";
RL Biochimie 70:1137-1142(1988).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).

QY 114 RASASY-YEYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKN 172
 DB 112 RSTNTFNATYHTDEIYDFMDLLVAEHPQLVSKLQIGRSYEGRPYLYLVKFS-TGGSNRP 170
 QY 173 AIWIDCGIHAREWISPAFLWFI----- 195
 DB 171 AIWIDLGHSRWITQATGVFAKFTEDYQDPSFTALDSMDIFLEIVTNPDCGFAETH 230
 QY 196 GHRMWRKNSRYANNHCIGTDLNSFNVSKHWCEGASSSCSYCYGLYPESEPEVKAV 255
 DB 231 SQNLWRKTRSYTSSSLCGVDPNRNW-DAGFGKAGASSPCSETYHGKYANSEVEVKSI 289
 QY 256 ASFLRNINOIKAYISMSYSHIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKTR 315
 DB 290 VDFVXDHGN-FKAFLSIHSYQLLYPYGYTTQSPDKTELNQVAKSAVEAL-KSLYGTG 347
 QY 316 YTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346
 DB 348 YKYGSIITTYQASGSDWSYNQGIKYSFT 378

RESULT 11
 ID_CBP1_RAT STANDARD; PRT; 419 AA.
 AC P00731;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPA1
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82105986; PubMed=6275388;
 RA Quinto C., Quiraga M., Swain W.F., Nikovits W.C. Jr., Standing D.N.,
 RA Pictet R.L., Valenzuela P., Rutter W.J.;
 RT "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
 RT characterization of the gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89034324; PubMed=3182872;
 RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
 RT "Structural characterization of the rat carboxypeptidase A1 and B
 RT genes. Comparative analysis of the rat carboxypeptidase gene
 RT family";
 RL J. Biol. Chem. 263:17837-17845(1988).
 CC -I- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
 CC amino acid.
 CC -I- SUBUNIT: Monomer.
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; V01232; CAA24542.1; -
 CC EMBL; J00713; AAA40893.1; -
 CC EMBL; M23990; AAA40955.1; -
 CC EMBL; M23960; AAA40955.1; JOINED.
 CC EMBL; M23985; AAA40955.1; JOINED.
 CC EMBL; M23986; AAA40955.1; JOINED.
 CC EMBL; M23987; AAA40955.1; JOINED.
 CC EMBL; M23988; AAA40955.1; JOINED.
 CC EMBL; M23989; AAA40955.1; JOINED.

PIR: A00911; CPRTA.
 DR HSSP; P00730; 1PYT.
 DR MEROPS; M4.001; -
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn carboPept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn carboPept; 1.
 DR SMART; SM00631; Zn_pept; 1.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
 FT SIGNAL 1 16
 FT PROPEP 17 110 ACTIVATION PEPTIDE.
 FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
 FT METAL 179 179 ZINC (BY SIMILARITY).
 FT METAL 182 182 ZINC (BY SIMILARITY).
 FT METAL 306 306 ZINC (BY SIMILARITY).
 FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE -380 380 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 248 271 BY SIMILARITY.
 FT CONFLICT 196 196 I -> V (IN REF. 1).
 FT CONFLICT 261 263 FGM -> LGK (IN REF. 1).
 FT CONFLICT 347 347 K -> E (IN REF. 1).
 SQ SEQUENCE 419 AA; 47197 MW; BB002D1CB99B7491 CRC64;

Query Match 27.7%; Score 529.5; DB 1; Length 419;
 Best Local Similarity 34.5%; Pred. No. 2.4e-37;
 Matches 134; Conservative 63; Mismatches 136; Indels 55; Gaps 10;

QY 2 KLCSLAVLVPIVLFCQHFVAFQSGQVLAALPRTSRQVQLNLTITTYEIVL--WQPVTA 59
 DB 3 RLLTSLLEAVCGNEN---FVGHQVLRISAADQAQVKVELEDLEHLQDFFWRDAAR 58
 QY 60 DLIVKKQKVHFFVNASDVNDVKAHLNLSVGIKPCSVLLADVEDLI---QQQISNDTVSPRAS 116
 DB 59 AGIPIDVRVPF---PSIQSVKAFLEYHGISVEIMEDVQLLDEKQMSAFQARALST 114
 QY 117 ASY-YEYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKN 175
 DB 115 DSFNATYHTDEIYDFMDLLVAEHPQLVSKLQIGRSYEGRPYLYLVKFS-TGGINRPALV 173
 QY 176 IDCQIHAREWISPAFLWFI-----N 198
 DB 174 IDTGHSRWITQATGVFAKFTEDYQDPSFTALDSMDIFLEIVTNPDCGFAETHKN 233
 QY 199 RWRMWRKNSRYANNHCIGTDLNSFNVSKHWCEGASSSCSYCYGLYPESEPEVKAVASF 258
 DB 234 RWRMWRKNSRYANNHCIGTDLNSFNVSKHWCEGASSSCSYCYGLYPESEPEVKAVASF 292
 QY 259 LRNNINOIKAYISMSYSHIVFPYSTRSKSDHEELSLVASEAVRAIDKTSKTR 318
 DB 293 VTSHGN-IFAFISHSYQLLYPYGYTSEFAPDQAEIDQLAKSAVTAL-TSLHGTKPKY 350
 QY 319 GHGSETLYLAPGGDDWIYDLGIKYSFT 346
 DB 351 GSIIDTIYQASGSDWSYNQGIKYSFT 378

RESULT 12
 ID_CBP2_HUMAN STANDARD; PRT; 417 AA.
 AC P48052; Q96A12; Q96QN3;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
 GN CPA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.

RC TISSUE=Pancreas;
RX MEDLINE=95204457; PubMed=7896805;
RA Cateaus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
RA Billeter M.;
RT "The sequence and conformation of human pancreatic
RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT three-dimensional model.";
RL J. Biol. Chem. 270:6651-6657(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain, and Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins E.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski S.J., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98111000; PubMed=9450539;
RA Reverter D., Garcia-Saez I., Cateaus L., Vendrell J., Coll M.,
RA Aviles F.X.;
RT "Characterisation and preliminary X-ray diffraction analysis of human
RT pancreatic procarboxypeptidase A2.";
RL FEBS Lett. 420:7-10(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=98046021; PubMed=9384570;
RA Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT "The three-dimensional structure of human procarboxypeptidase A2.
RT Deciphering the basis of the inhibition, activation and intrinsic
RT activity of the zymogen.";
RL EMBO J. 16:6906-6913(1997).
CC -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC residues.
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U19977; AAA74425.1; -;
CC EMBL; BC007009; AAH07009.1; -;
CC EMBL; BC014571; AAH14571.1; -;
CC EMBL; BC015140; AAH15140.1; -;
CC PIR; A56171; A56171.
CC PDB; 1AYE; 13-JAN-99.
CC PDB; 1D7D; 12-JUL-00.
CC PDB; 106X; 30-JAN-03.
CC MEROPS; M14.002; -;
CC Genew; HGNC:2297; CPA2.
CC MIM; 600688; -;
CC GO; GO:0004182; F:carboxypeptidase A activity; TAS.

DR GO; GO:0007039; P:vacuolar protein catabolism; TAS.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 16 POTENTIAL.
FT PROPEP 17 112 ACTIVATION PEPTIDE.
FT CHAIN 113 417 CARBOXYPEPTIDASE A2.
FT METAL 177 177 ZINC.
FT METAL 180 180 ZINC.
FT METAL 304 304 ZINC.
FT ACT_SITE 378 378 NUCLEOPHILE.
FT DISULFID 246 269
FT DISULFID 318 352
FT CONFLICT 80 80
FT CONFLICT 302 302 E -> G (IN REF. 2; AAH07009).
FT CONFLICT 302 302 T -> I (IN REF. 1).
SQ SEQUENCE 417 AA; 46828 MW; 3D9C073A98700A83 CRC64;
Query Match 26.7%; Score 511; DB 1; Length 417;
Best Local Similarity 32.5%; Pred. No. 8.9e-36;
Matches 128; Conservative 64; Mismatches 136; Indels 66; Gaps 11;
QY 19 HVF---AFQSGVLAALPRTSRQVQLQNLTITYIVL--WQPVTADLVKKQVHFFVN 73
DB 13 HIYCLTFVGDQVLEIVPSNEEQIKNLQLQAEHLQDLDFWKSPT----TPGETAHVRVP 68
QY 74 ASDVDNVAHLNVSGIPGCVLLADVEDLIQQQISNDTVSPRASAS---YYEQYHSLNETY 130
DB 69 FVNVQAVKVFELSGQIAVSIMEDVQVLLDRENEMLFNRRERSGNFNGAYHTLEELS 128
QY 131 SWIEPITERHDPMLTKIHIGSFKEPKYPLYLVKVS--GKEQTAKNAIWDGCIHARWISPA 189
DB 129 QEMDNLVAEHPLGVLKVNIGSFENRPMNVLFKFTSGD---KPAIWLDAIGHAREWVQA 185
QY 190 FCLWFIG-----HNRWRKRSFYANNH 212
DB 186 TALWTANKIVSDYKDPISITILDALDIFLLPVTNPDGYVFSQTKNRWRKTRSKVSGSL 245
QY 213 CIGTDLNSNFVSKHWCEEGASSSCSETYCGLYPESEPEKAVASFLRNINQIKAYISM 272
DB 246 CVGVDPNRNW--DAGFGGFGASSNCPDSYHGPSANSEVEKSIYDFIKSH-GKVKAFITL 303
QY 273 HSYQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGG 332
DB 304 HSYQLLMFPYGYKTKLDDFDLSEVAQKAAQSL-RSLHGTKYKVGPICSVIYQASGGS 362
QY 333 DDWLYDLGIKYSFTSN-----PPVEKLIP 356
DB 363 IDWSYDYGIKYSFAFELRDTGRYGLLPARQILP 396
RESULT 13
CBP1_HUMAN
ID CBP1_HUMAN STANDARD; PRT; 419 AA.
AC P15085; Q9BS67;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN CPA1 OR CPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;

RX MEDLINE=93038569; PubMed=1417781;
RA Cateasus L., Villegas V., Pascual R., Aviles F.X.,
RA Wicker-Planquart C., Puigserver A.,
RT "cDNA cloning and sequence analysis of human pancreatic
RT procarboxypeptidase A1";
RL Biochem. J. 287:299-303(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 17-42.
RX MEDLINE=89153096; PubMed=2920728;
RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT "Purification and properties of five different forms of human
RT procarboxypeptidases.";
RL Eur. J. Biochem. 179:609-616(1989).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X67318; CAA47732.1; -.
DR EMBL; BC005279; AAH05279.1; -.
DR PIR; S29127; S29127.
DR HSP; P00730; 1PVT.
DR MEROPS; M14.001; -.
DR Genew; HGNC:2296; CPA1.
DR MIM; 114850; -.
DR GO; GO:0004182; P:carboxypeptidase A activity; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carboxypeptidase.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypeptidase; 1.
DR PRINTS; PR00765; CARBOXYPTASA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 182 182 ZINC (BY SIMILARITY).

FT METAL 306 306 ZINC (BY SIMILARITY).
FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 248 271 BY SIMILARITY.
FT CONFLICT 276 276 H -> R (IN REF. 2).
SQ SEQUENCE 419 AA; 47140 MW; 439FAFFFAEE958B1 CRC64;

Query Match 25.6%; Score 489.5; DB 1; Length 419;
Best Local Similarity 32.7%; Pred. No. 5.9e-34;
Matches 131; Conservative 60; Mismatches 12; Indels 83; Gaps 13;

QY 3 LCSLAVLVPIVLFCQHQVFAFGQGVLAALPRTRSRQVQLQNTLTYYEIVL--WQ----- 55
Db 4 LLVLSVLGAVFGKED-----FVGHQVLRAISVADAEAQVKVKELEDLEHLQDLDFWGRGPAHP 59

QY 56 --PVTADLIVKKQVHFVNASDVNDVNAHLNVGIPCSVLLADVEDLI---QQQI----- 106
Db 60 GSPIDV-----RVPF-----PSIQAVKIFLESIGSYETMTEDVQSLLDEEQEQMFAPR 108

QY 107 ---SNDTVSPRASASYVEQYHSLNEIYSWIEFTETHPDMLTKIHGSSPEKYPYLVK 162
Db 109 SRARSTDTFN-----YATHTLEIYDFDLVAENPHLVSKIQIGNTYEGREIYVLK 161

QY 163 VSGKEQTAKNAIWDICGHAREWISPAFLWFI----- 195
Db 162 FS-TGSKRPALWDTGHSREWVTQASGVWFACKITODYGQDAAFATLDTLDFLEIV 220

QY 196 -----GHNMRWRKNSFYANNHCIGTDLNSNFVSKWCERGASSSCSETYCGLY 245
Db 221 TNPDGFAFTHSTNRMRKTRSHTAGSLCIGVDPNRNW-DAGFGLSGASSNPCSETYHGKF 279

QY 246 PESEPEVKAVASFLRRNLNQIKAYISMHSYSHQIVFPYSYTRSKSKDHEELSLVASEAVR 305
Db 280 ANSEVEKSIYDFVKDHGN-IRAFISHSYQLMYPYGYKTEPVDPDQDLQLSKAAVT 338

QY 306 AIDTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346
Db 339 AL-ASLYGKFNYSIIKAIYQASGSTIDWYSQGIKYSFT 378

RESULT 14
CBP4 HUMAN
ID CBP4 HUMAN STANDARD; PRT; 421 AA.
AC Q9UJ42;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN CPA4 OR CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99310537; PubMed=10383164;
RA Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.L.;
RT "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT deacetylase inhibitors during differentiation of prostate epithelial
RT cancer cells.";
RL Cancer Res. 59:2981-2988(1999).
RN [2]
RP IMPRINTING.
RX MEDLINE=224483629; PubMed=12552318;
RA Kayashima T., Yamasaki K., Yamada T., Sakai H., Miwa N., Ohta T.,
RA Yoshiura K.-I., Matsumoto N., Nakane Y., Kanetake H., Ishino F.,
RA Nikawa N., Kishino T.;
RT "The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32
RT imprinting domain.";
RL Hum. Genet. 112:220-226(2003).
CC -!- FUNCTION: Could be involved in the histone hyperacetylation
CC pathway.
CC -!- TISSUE SPECIFICITY: Fetal expression in the adrenal gland, brain,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:19:56 ; Search time 60.8768 Seconds
(without alignments)
1526.015 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLPIVLFCEQHV.....IKYFTSNPPVKKLLPLSLK 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rhodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1884	98.6	360	Q9P2Y6	Q9P2Y6 homo sapien
2	1799.5	94.2	423	Q15114	Q15114 homo sapien
3	1794.5	93.9	423	Q961Y4	Q961Y4 mus musculus
4	1528	80.0	422	Q9JHH6	Q9JHH6 mus musculus
5	1522	79.6	422	Q9QZFO	Q9QZFO mus musculus
6	1499	78.4	422	Q9EQV9	Q9EQV9 rattus norv
7	1029	53.8	198	Q9NTI8	Q9NTI8 homo sapien
8	700.5	36.7	416	Q9PUP2	Q9PUP2 bothrops ja
9	682	33.7	408	Q9AXN3	Q9AXN3 paralicthy
10	646	33.8	416	Q9XSP3	Q9XSP3 sus scrofa
11	621	32.5	412	Q97597	Q97597 rattus norv
12	606	31.7	434	Q9AVS2	Q9AVS2 xenopus lae
13	602	31.5	437	Q8N4T0	Q8N4T0 homo sapien
14	581	30.4	419	Q8UUK1	Q8UUK1 gallus gall
15	563	29.5	374	Q8IVL8	Q8IVL8 homo sapien
16	554.5	29.0	436	Q8R4H4	Q8R4H4 mus musculus

17	529.5	27.7	419	13	Q8AXN4	Q8AXN4 paralicthy
18	519.5	27.2	436	4	Q8WXQ8	Q8WXQ8 homo sapien
19	516.5	27.0	419	6	Q9TV85	Q9TV85 sus scrofa
20	506.5	26.5	420	11	Q8BMK6	Q8BMK6 mus musculus
21	484.5	25.4	452	13	Q8UW45	Q8UW45 fugu rubrip
22	477.5	25.0	262	11	Q8BVD0	Q8BVD0 mus musculus
23	471.5	24.7	247	4	Q86K29	Q86K29 homo sapien
24	460.5	24.1	419	13	Q8AXN5	Q8AXN5 paralicthy
25	460	24.1	424	5	Q9VL87	Q9VL87 drosophila
26	460	24.1	428	5	Q8IGC4	Q8IGC4 drosophila
27	459	24.0	1132	5	Q9W475	Q9W475 drosophila
28	451.5	23.6	422	5	Q9VLZ2	Q9VLZ2 drosophila
29	440.5	23.1	226	4	Q8TDE8	Q8TDE8 homo sapien
30	436	22.8	584	5	Q9TZH1	Q9TZH1 caenorhabdi
31	434.5	22.7	351	5	Q8TOQ1	Q8TOQ1 drosophila
32	434	22.7	351	4	Q8NA08	Q8NA08 homo sapien
33	418	21.9	445	5	Q9W478	Q9W478 drosophila
34	416	21.8	488	5	Q9BL98	Q9BL98 caenorhabdi
35	412.5	21.6	416	5	Q9VRZ3	Q9VRZ3 drosophila
36	407.5	21.3	419	5	Q9VLZ1	Q9VLZ1 drosophila
37	402	21.0	424	5	Q9VS66	Q9VS66 drosophila
38	401.5	21.0	430	5	Q8SZP6	Q8SZP6 drosophila
39	400.5	21.0	430	5	Q9VL86	Q9VL86 drosophila
40	399.5	20.9	427	5	Q9U9K2	Q9U9K2 aedes aegypt
41	397	20.8	323	5	Q19121	Q19121 caenorhabdi
42	395.5	20.7	540	5	Q9TZH2	Q9TZH2 caenorhabdi
43	391	20.5	453	5	Q9V342	Q9V342 drosophila
44	385	20.1	423	5	Q61532	Q61532 drosophila
45	382.5	20.0	581	5	Q23318	Q23318 caenorhabdi

ALIGNMENTS

RESULT 1

Q9P2Y6 ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Carboxypeptidase B-like protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
carboxypeptidase B from human plasma."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
carboxypeptidase B from human plasma."
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; AB011969; BAA90475.1; -;
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -;
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carboOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 360 AA; 40915 MW; BF670B2F7437C1CB CRC64;

Query Match 98.6%; Score 1884; DB 4; Length 360;
Best Local Similarity 98.3%; Pred. No. 4.9e-162;
Matches 354; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
QY 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTVEIWLQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTVEIWLQPVTD 60
QY 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLLQQQISNDTSPRASASY 120
DB 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLLQQQISNDTSPRASASY 120
QY 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKPYLVLVKVSKEQAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKPYLVLVKVSKEQAKNAIWDGCI 180
QY 181 HAREMISPACFLWFIQHNRWKRNSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSET 240
DB 181 HAREMISPACFLWFIQHNRWKRNSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSET 240
QY 241 YCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSQHIYFFYSYTRSKSKDHEELSLVA 300
DB 241 YCGLYPESEPEVKAVASFLRRNIQIKAYISMHSYSQHIYFFYSYTRSKSKDHEELSLVA 300
QY 301 SEAVRAIDKTSKNTRYTHGHGSETIYLPAGGGDDWIYDLGIKYSFTSPNPPVEKLLPLSLK 360
DB 301 SEAVRAIEKISKNTRYTHGHGSETIYLPAGGGDDWIYDLGIKYSFTSPNPPVEKLLPLSLK 360

RESULT 2
Q15114
ID Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PCPB protein.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC TISSUE=Liver;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Teal S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14_009; -.
DR Genew; HGNC:2300; CPB2.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn carboxypept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPEPTASEA.
DR SMART; SM00631; Zn_pept; 1.
FT CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6EE79F4 CRC64;

Query Match 94.2%; Score 1799.5; DB 4; Length 423;
Best Local Similarity 89.6%; Pred. No. 2.7e-154;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTVEIWLQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTVEIWLQPVTD 60
QY 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLLQQQISNDTSPRASASY 120
DB 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLLQQQISNDTSPRASASY 120
QY 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKPYLVLVKVSKEQAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKPYLVLVKVSKEQAKNAIWDGCI 180

Query Match 94.2%; Score 1799.5; DB 4; Length 423;
Best Local Similarity 89.6%; Pred. No. 2.7e-154;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;
```

RESULT 3

```
Q961Y4
ID Q961Y4 PRELIMINARY; PRT; 423 AA.
AC Q961Y4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B2 (Plasma).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007057; AAH07057.1; -.
DR MEROPS; M14_009; -.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn carboxypept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPEPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 423 AA; 48412 MW; 9B383B03F6EE7CF5 CRC64;

Query Match 93.9%; Score 1794.5; DB 4; Length 423;
Best Local Similarity 89.3%; Pred. No. 7.7e-154;
Matches 342; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

QY 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTVEIWLQPVTD 60
DB 1 MKLCSLAVLPVILFCEQHVFAFQSGQVLAALPRTSRQVQLQNLTTTVEIWLQPVTD 60
QY 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLLQQQISNDTSPRASASY 120
DB 61 LIVKKQVHFFVNASDVNDVNVKAHLNVSGIPCSVLLADVEDLLQQQISNDTSPRASASY 120
QY 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKPYLVLVKVSKEQAKNAIWDGCI 180
DB 121 EOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKPYLVLVKVSKEQAKNAIWDGCI 180
QY 181 HAREMISPACFLWFIQHNRWKRNSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSET 203
DB 181 HAREMISPACFLWFIQHNRWKRNSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSET 240
QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
DB 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300
QY 264 NOIKAYISMHSYSQHIYFFYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
DB 264 NOIKAYISMHSYSQHIYFFYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 360

Query Match 93.9%; Score 1794.5; DB 4; Length 423;
Best Local Similarity 89.3%; Pred. No. 7.7e-154;
Matches 342; Conservative 1; Mismatches 3; Indels 37; Gaps 1;
```

DB 301 NQIKAYISMHSYSHVFPYSTRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGSGE 360

QY 324 TLYLAPGGDDWIYDLGIKYSFT 346

DB 361 TLYLAPGGDDWIYDLGIKYSFT 383

RESULT 4

Q9JHH6 ID Q9JHH6 PRELIMINARY; PRT; 422 AA.

AC Q9JHH6; (1)

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor) (110032P04RIK protein)

DE CPB2 OR TAPI OR 1110032P04RIK.

GN Mus musculus (Mouse)

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NCBI_TaxID=10090;

OX (1)

RN SEQUENCE FROM N.A.

RP MEDLINE=20341711; PubMed=10878383;

RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N., Campbell W., Okada H.;

RA "pro-carboxypeptidase R is an acute phase protein in the mouse, whereas carboxypeptidase N is not.;"

RL J. Immunol. 165:1053-1058(2000).

RN (2)

RP SEQUENCE FROM N.A.

RX MEDLINE=20201996; PubMed=10739389;

RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J., van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;

RA "Characterization of mouse thrombin-activatable fibrinolysis inhibitor.;"

RL Thromb. Haemost. 83:297-303(2000).

RN (3)

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Embryo;

RX MEDLINE=21085660; PubMed=11217651;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kado K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Hayashizaki Y.;

RA "Functional annotation of a full-length mouse cDNA collection.;"

RL Nature 409:685-690(2001).

DR EMBL; AB021968; BAB03402.1; -;

DR EMBL; AF164524; BAF62385.1; -;

DR EMBL; AK004045; BAB23141.1; -;

DR HSSP; P00730; 2CTC.

DR MEROPS; M14.009; -;

DR MGD; MGI:1891837; Cpb2.

DR InterPro; IPR003146; Propep M14.

DR Pfam; PF02244; Propep M14; 1.

DR Pfam; PF00246; Zn carboxypeptidase 1.

DR PRINTS; PR00765; CRBOXYPBASEA.

DR SMART; SM00631; Zn_pept; 1.

SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

Query Match 80.0%; Score 1528; DB 11; Length 422;

Best Local Similarity 76.2%; Pred. No. 9.9e-130;

Matches 292; Conservative 24; Mismatches 29; Indels 38; Gaps 2;

QY 1 MKLCSLAVLVPIVLFCEQHVFAFOSGVLAALPRTSRQVQLNLTITTYEIVLWQPVTAAD 60

DB 1 MKLHGLGLVAILLY-EQGFQFOSGVLSALPRTSRQVQLNLTITTYEIVLWQPVTAAD 59

QY 61 LIVKKQVHFFVNASDVNDVNAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASASY 120

DB 60 FIEKKEVHFFVNASDVDSVAHLNVSRIPFNVLNNVEDLIEQTFNDTVSPRASASY 119

QY 121 EYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKLYPLVKVSGKEQAKNAIWDCCI 180

DB 120 EYHSLNEIYSWIEVITEHQPDMLQKIYIGSSFYKLYPLVKVSGKEQAKNAIWDCCI 179

QY 181 HAREWISPAFLWIGH-----NRMRK 203

DB 180 HAREWISPAFLWIFGYVTFPHGKENLYTRLLRHVDYFIMPVNVVDYDTWKQRMWRK 239

QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSETYCGLYPSEPEVKAVASFLRNNI 263

DB 240 NRSFAKNRRCVGTDLNRFNFAKHWCEKASSSSCSETYCGLYPSEPEVKAVADFLRNNI 299

QY 264 NQIKAYISMHSYSHVFPYSTRSKSKDHELSLVASEAVRAIDKTSKNTRYTHGSGE 323

DB 300 DHIKAYISMHSYSHVFPYSTRSKSKDHELSLVASEAVRAIESINKNTRYTHGSGE 359

QY 324 TLYLAPGGDDWIYDLGIKYSFT 346

DB 360 TLYLAPGGDDWIYDLGIKYSFT 382

RESULT 5

Q9QZF0 ID Q9QZF0 PRELIMINARY; PRT; 422 AA.

AC Q9QZF0;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Carboxypeptidase U.

GN CPB2.

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NCBI_TaxID=10090;

OX (1)

RN SEQUENCE FROM N.A.

RC STRAIN=B6/CBAF1J;

RA He Y.C., Broze G.;

RT "Isolation and characterization of mouse liver carboxypeptidase B gene.;"

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF186188; AAF00528.1; -;

DR HSSP; P00730; 5CPA.

DR MEROPS; M14.009; -;

DR MGD; MGI:1891837; Cpb2.

DR InterPro; IPR003146; Propep M14.

DR Pfam; PF02244; Propep M14; 1.

DR Pfam; PF00246; Zn carboxypeptidase 1.

DR PRINTS; PR00765; CRBOXYPBASEA.

DR SMART; SM00631; Zn_pept; 1.

SQ SEQUENCE 422 AA; 48832 MW; B53FFPFB09943954E CRC64;

Query Match 79.6%; Score 1522; DB 11; Length 422;

Best Local Similarity 76.0%; Pred. No. 3.5e-129;

Matches 291; Conservative 24; Mismatches 30; Indels 38; Gaps 2;

QY 1 MKLCSLAVLVPIVLFCEQHVFAFOSGVLAALPRTSRQVQLNLTITTYEIVLWQPVTAAD 60

DB 1 MKLHGLGLVAILLY-EQGFQFOSGVLSALPRTSRQVQLNLTITTYEIVLWQPVTAAD 59

```
Db 1 MKLHGLGILVAIILY-EQHGFAFQSGQVLSALPRTSRQVQLLQNLTTTVEVVLWQPVTAE 59
QY 61 LIVKKQVHFFVNASDVNKAHLNVSGIPCSVLADVEDLIQQOISNDTSPRASASY 120
Db 60 FIEKKVEHFFVNASDVSKAHLNVSRIPFVLMNVEDLIEQQTNDTSPRASASY 119
QY 121 EYHSLNEIYSWIEFTEHRHPDMLTKIHGSSFEKYPYLVKVSKEQOTAKNAIWDGCI 180
Db 120 EYHSLNEIYSWIEVTEHQDPMQLKIYIGSSFEKYPYLVKVSKEQRIKNAIWDGCI 179
QY 181 HAREWISPAFCILWFIGH-----NRMWRK 203
Db 180 HAREWISPAFCILWFIGYVTFPHGKENTYTRLLRHVDYFIMPVNMVDGYDTWKQNRMRK 239
QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRNI 263
Db 240 NRSFHNKNNRCVGTDLNRNFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVADFLRNI 299
QY 264 NOIKAYISMHSYQIHTVPYSTRSKSKDHELSLVASEAVRAIDKTSKNTRYTHGSGE 323
Db 300 DHIKAYISMHSYQILFPYSYNRSKSKDHELSLVASEAVRAIGSFNKNTRYTHGSGE 359
QY 324 TLYLAPGGDDWIYDLGIKYSFT 346
Db 360 SLYLAPGGSDWIYDLGIKYSFT 382
RESULT 6
Q9EQV9 PRELIMINARY; PRT; 422 AA.
AC Q9EQV9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Pre-procarboxypeptidase R.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA Okada N., Okada H.; and partial characterization of rat
RT "Molecular cloning and partial characterization of rat
RT procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728(2000).
DR EMBL; AB042598; BAB18617.1; -.
DR HSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
KW Carboxypeptidase.
SQ SEQUENCE 422 AA; 48826 MW; FFFD32A51A9366C8 CRC64;
Query Match 78.4%; Score 1499; DB 11; Length 422;
Best Local Similarity 74.7%; Pred. No. 4.2e-127;
Matches 286; Conservative 26; Mismatches 33; Indels 38; Gaps 2;
QY 1 MKLCSLAVLVPILFCEQHVFAFQSGQVLAALPRTSRQVQLNLTTTVEVVLWQPVTAD 60
Db 1 MKLYGLGLVAIILY-EKHGLAFQSGHVLASLPRTSRQVQLLQNLTTTVEVVLWQPVTAE 59
QY 61 LIVKKQVHFFVNASDVNKAHLNVSGIPCSVLADVEDLIQQOISNDTSPRASASY 120
Db 60 FIEKKVEHFFVNASDVSKAHLNVSRIPFVLMNVEDLIEQQTNDTSPRASASY 119
QY 121 EYHSLNEIYSWIEFTEHRHPDMLTKIHGSSFEKYPYLVKVSKEQOTAKNAIWDGCI 180
Db 120 EYHSLNEIYSWIEVTEHQDPMQLKIYIGSSFEKYPYLVKVSKEQRIKNAIWDGCI 179
```

```
QY 181 HAREWISPAFCILWFIGH-----NRMWRK 203
Db 180 HAREWISPAFCILWFIGYVTFPHGKENTYTRLLRHVDYFIMPVNMVDGYDTWKQNRMRK 239
QY 204 NRSFYANNHCIGTDLNSNFVSKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRNI 263
Db 240 NRSFHNKNNRCVGTDLNRNFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVADFLRNI 299
QY 264 NOIKAYISMHSYQIHTVPYSTRSKSKDHELSLVASEAVRAIDKTSKNTRYTHGSGE 323
Db 300 NHIKAYISMHSYQILFPYSYNRSKSKDHELSLVASEAVRAIESINKNTRYTHGSGE 359
QY 324 TLYLAPGGDDWIYDLGIKYSFT 346
Db 360 SLYLAPGGSDWIYDLGIKYSFT 382
RESULT 7
Q9NTI8 PRELIMINARY; PRT; 198 AA.
AC Q9NTI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BA139H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RL Tromans A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137141; CAB92622.1; -.
DR HSP; P00730; 1CPX.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
FT NON_TER 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;
Query Match 53.8%; Score 1029; DB 4; Length 198;
Best Local Similarity 99.5%; Pred. No. 4.9e-85;
Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MKLCSLAVLVPILFCEQHVFAFQSGQVLAALPRTSRQVQLNLTTTVEVVLWQPVTAD 60
Db 1 MKLCSLAVLVPILFCEQHVFAFQSGQVLAALPRTSRQVQLNLTTTVEVVLWQPVTAD 60
QY 61 LIVKKQVHFFVNASDVNKAHLNVSGIPCSVLADVEDLIQQOISNDTSPRASASY 120
Db 61 LIVKKQVHFFVNASDVNKAHLNVSGIPCSVLADVEDLIQQOISNDTSPRASASY 120
QY 121 EYHSLNEIYSWIEFTEHRHPDMLTKIHGSSFEKYPYLVKVSKEQOTAKNAIWDGCI 180
Db 121 EYHSLNEIYSWIEFTEHRHPDMLTKIHGSSFEKYPYLVKVSKEQAAKNAIWDGCI 180
QY 181 HAREWISPAFCILWFIGH 197
Db 181 HAREWISPAFCILWFIGH 197
RESULT 8
Q9PUF2 PRELIMINARY; PRT; 416 AA.
AC Q9PUF2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
```

```
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase homolog.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodermata; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Vipidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT "Screening of Bothrops jararaca pancreas cDNA library.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF190274; AAF01344.1; -.
DR HSP; P09955; INSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypeptidase.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypeptidase; 1.
DR PRINTS; P00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;

Query Match 36.7%; Score 700.5; DB 13; Length 416;
Best Local Similarity 38.5%; Pred. No. 7.3e-55;
Matches 142; Conservative 65; Mismatches 119; Indels 43; Gaps 5;

QY 15 FCEQVFAFQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPVTDADLVKKKQVHFFVNA 74
DB 14 FAETTVHRPDGKVRVTRPNEDEYFLNLANIVQVDFWPRDSVELKAEKMTVDRIEA 73
QY 75 SDVDNVKALHNSGIPCSVLADVEDLIQQOISNDTVSPRASASYEYQVHSLNEIYSWIE 134
DB 74 DECSEVESLQSGNLNLEIDNLQAVLDRLQDNIA---RTAGYNEKINSWEKIDAWTA 130
QY 135 FTERHPDMLTKIHIGSSFEKYPVLYLVKSGKEQAKNAIWDICGHHAREWISPAFLCWF 194
DB 131 DIANENPSLVSLQIGTTTFEGRPMLLV-KKPGVKKKAIKFDICGHHAREWISPAFLCWF 189
QY 195 I-----GHNRMWRKNSFYANNHICIGTD 217
DB 190 VREAVTYGKETIMTQLNKLDPYILPVNLIDGYVSWKQSRMRKTRSNAGSTCIGTD 249
QY 218 LNSNFVSHWCEGASSSCSTYCGLYPESPEVKAVASFLRRNINQIKAYISMHSYSQ 277
DB 250 PNRNF-DAACVSGASRPPCSTYCGSPSEKTKALADFIKRRNSIIQAVLIHSYSQ 308
QY 278 HIVFPYSYTRSKDHELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDMIY 337
DB 309 MLLYPYSYTYDUTSNKKLNSIAKEAIREL-KVLPGETYTYGGAATYPAAGGSDWAY 367
QY 338 DLGIKYSPT 346
DB 368 DQGIKYAFT 376

RESULT 9
Q8AXN3 PRELIMINARY; PRT; 408 AA.
AC Q8AXN3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Carboxypeptidase B (fragment).
GN CPB.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralicththyidae; Paralicthys.

OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RA Srivastava A.S., Kurokawa T., Suzuki T.;
RT "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
A2 and B from the Japanese flounder, Paralicthys olivaceus.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB099302; BAC53789.1; -.
FT NON TER 1
SQ SEQUENCE 408 AA; 45904 MW; 809CAACD9D2DA34B CRC64;

Query Match 35.7%; Score 682; DB 13; Length 408;
Best Local Similarity 39.3%; Pred. No. 3.4e-53;
Matches 142; Conservative 61; Mismatches 116; Indels 42; Gaps 5;

QY 23 FQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPVTDADLVKKKQVHFFVNASDVDNVKA 82
DB 12 FEGEKVPFLKPVFDEHVTILRLANSIEVDWPRSSQVLTIDIDVDIRVPAIYLDWVYT 71
QY 83 HLNVSGLPCSVLLADVEDLIQQOISNDTVSPRASASYEYQVHSLNEIYSWIEFITERHPD 142
DB 72 LLOQSDMEHEVLIEDLSAVDAEAD---LKPSRAHSYTKYNLDKQVSWIASISSNPD 128
QY 143 MLTKIHIGSSFEKYPVLYLVKSGKEQAKNAIWDICGHHAREWISPAFLCWF----- 195
DB 129 LISQVIGNTYEGRPMTVLKLGKSSSTKPAIFMDCGHHAREWISPAFCQFVKEALSTY 188
QY 196 -----GH-----NRMWRKNSFYANNHICIGTDLNSNFVSK 225
DB 189 GSDSQMTSLNQMDVFLPVENIDGYDFTKSNRMWRKTRSKSGSSCIGADPNRNF-DA 247
QY 226 HWCSEGASSSCSTYCGLYPESPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
DB 248 GWCSGASDNPCSDTFCGYTPESIEVKNVADFIKRRNSIIKAYLVTHYSQGLLPYSY 307
QY 286 TRSKDHELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDMIYDLGIKYSF 345
DB 308 TYQAAADHSELKVAAGSAAAL-RSLYGTITYTSGPATTIYPAGGSDWAYDLGVKYSY 366
QY 346 T 346
DB 367 T 367

RESULT 10
Q9XSP3 PRELIMINARY; PRT; 416 AA.
AC Q9XSP3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Procarboxypeptidase B precursor (BC 3.4.17.2).
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=99321926; PubMed=10391940;
RA Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RT Hershberger C., Aviles F.;
RT "Mapping the Pro-region of carboxypeptidase B by protein engineering.
Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL J. Biol. Chem. 274:19925-19933(1999).
DR EMBL; AJ133775; CAB46991.1; -.
DR HSP; P09955; 1PEA.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypeptidase.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypeptidase; 1.
DR PRINTS; P00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
```



```
Db 184 KAIWIDCGHAREWIGPACQWFVKEAINSYNTDPMKILNLLIYMPVFNVDGHSYS 243
QY 198 ---NRMWRKNRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSTYCYGLYPESEPEVKA 254
Db 244 WHSDRFWRKTRSKNTRYCYGVDAARNW-KVHWSDEGASLNPDCNTYCGPYAESEPEVKA 302
QY 255 VASFLRNINQIKAYISMHSYSHQHVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNT 314
Db 303 VAQFLYKORKHVRAYMSHAYAQMLLYPSYQYGAIPNFGCVESAHAHNAVLAI-RSAYGI 361
QY 315 RYTHGHGSETLYLAPGGDDWIYDLGIKYSF 345
Db 362 RYRHGPASTLTLYTSSGSSMDWAYNGIPYSY 392

RESULT 13
Q8N4T0
ID Q8N4T0 PRELIMINARY; PRT; 437 AA.
AC Q8N4T0;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Similar to carboxypeptidase A6.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033684; AAH33684.1; -.
DR Genew; HGNC:17245; CPA6.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept. 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase.
FT SIGNAL 1 17
FT CHAIN 113 419
SQ SEQUENCE 437 AA; 51024 MW; 9970680D429A0A98 CRC64;

Query Match 31.5%; Score 602; DB 4; Length 437;
Best Local Similarity 34.4%; Pred. No. 6.4e-46;
Matches 134; Conservative 67; Mismatches 135; Indels 54; Gaps 10;

QY 1 MKLC--SLAVLPVILFCEQHV--AFQSGQVLAALPRTSRQVQLQNLTTTYEIVLWQP 56
Db 15 LPCLWLFKLQP-----GSHLYNNRYAGDKVIRFPKTEEEAYALKKISYQLKVDLWQP 70
QY 57 VTADLIVK--KKQVHFNASVDNVKALNVSGIPCSVLLADVEDLIQQQISNDTVSPR 114
Db 71 SSISYVSGTVVDVHVPQNGSRA--LLAFLQGANIQYKVLIEDLQKLEKSSLTQRNR 128
QY 115 ASAS--YVEQVHSLNEISWIEFITERHPDMLTKIHGSSFEPKPLVYLKVGKEOTAKN 172
Db 129 RSLSGYNEVHSLSEIQWMMHNLKTHSGLIHMFISGRSYEGRCFLFKL-GRRSRLKR 187
QY 173 AIWIDCGIHAREWISPAFLWFI----- 195
Db 188 AVWIDCGIHAREWIGPACQWFVKEALLTYKSDPAMRWMLNHLFYIMPVFNVDGHSYS 247
QY 196 GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCCEGASSSCSTYCYGLYPESEPEVKA 255
Db 248 TNDRFWRKTRSRNSRFRGRGVDAARNWVKV-WCEGASMHPCDDTYCGFPPESEPEVKA 306
QY 256 ASFLARNINQIKAYISMHSYSHQHVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNT 315
Db 307 ANFLRKRHLKRLAYLSFAYIAQMLLYPSYKYKATIPNFRFCVESAAYKAVNAL-QSVYGR 365
QY 316 YTHGHGSETLYLAPGGDDWIYDLGIKYSF 345
```

```
Db 366 YRYGPASTLTLYVSSGSSMDWAYKNGIPYAF 395

RESULT 14
Q8UUK1
ID Q8UUK1 PRELIMINARY; PRT; 419 AA.
AC Q8UUK1;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Preprocarboxypeptidase A precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=domesticus; TISSUE=Pancreas;
RA Hasegawa S., Honda K., Hikami Y.;
RL "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A precursor.";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; X64539; CAA45837.1; -.
DR MEROPS; M14.001; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept. 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Signal; Carboxypeptidase.
FT SIGNAL 1 17
FT CHAIN 113 419
SQ SEQUENCE 419 AA; 47006 MW; 299C334274BCBC9C CRC64;

Query Match 30.4%; Score 581; DB 13; Length 419;
Best Local Similarity 35.3%; Pred. No. 4.8e-44;
Matches 133; Conservative 65; Mismatches 107; Indels 72; Gaps 9;

QY 23 FQSGQVLAALPRTSRQVQLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVND--- 79
Db 21 FVGHQVLRIVPSSDAELQKVQLQS-----LEHLQLDFMLSPRGIGNFPVD 65
QY 80 -----VKAHLNVSGIPCSVLLADVEDLI-----QQQISNDTVSPRASASY-YEQYH 124
Db 66 IRVPPPSLQPVKAHLEANGVPYSIMIEDVQALVLEQOMLRRRRFVPLSTSTFDYTSYH 125
QY 125 SLNEIYSWIEFITERHPDMLTKIHGSSFEPKPLVYLKVGKEOTAKNAIWDGCIHARE 184
Db 126 TLBEIYAFMDLVLAENPNLVSKLEIGRTTENRPIYVLKFS-KGGTNRPAIWDITGHSRE 184
QY 185 WISPAFLWFI-----GHNRMWRKNRSFYA 209
Db 185 WVTQSGIWFAKKIVQEQDEGLANILDQMDIFLEIVTNPDCGFATHTSNRMWRKTRSKRS 244
QY 210 NNHCIGTDLNSNFVSKHWCCEGASSSCSTYCYGLYPESEPEVKAFLARNINQIKAY 269
Db 245 GSLCGVGDPNRNW-DAGFGGSGASSNPCSETYHGPYANSEPEVKAIVDFVSKKN-IKAF 302
QY 270 ISMHSYSHQHVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAP 329
Db 303 ISIHYSQLLYPYGYTTTAVPDKEELHQAQVAAVAAL-SSLYGTNYKYSIITIIYQAS 361
QY 330 GGGDDWIYDLGIKYSFT 346
Db 362 GGTIDWTYNGIKYSFT 378

RESULT 15
```

Search completed: January 1, 2004, 19:31:55
Job time : 62.8768 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:33:22 ; Search time 393.069 Seconds
(without alignments)
2472.334 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSIAVLVPIVLCFQHV.....IKYSFTSNPPVKKLLPLSLK 360

Scoring table:
BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgm2_1/USPTO_spool/US09980881/runat_31122003_135914_26819/app_query.fasta_1.1628
-DB=N_Geneseq_19Jun03 -QPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MTN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09980881 -CGN 1 1 577 @runat_31122003_135914_26819 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1911	100.0	1573	21	AAC81962 Human brain carbox
2	1849.5	96.8	1625	25	ABX13670 Human protease CDN
3	1832	95.9	1272	20	AAV74302 Human plasma carbo
4	1832	95.9	1728	24	ABN95899 Gene #2397 used to
5	1832	95.9	1749	17	AAT11671 Human plasma carbo
6	1832	95.9	1749	18	AAT62846 Human plasma carbo
7	1827	95.6	1749	14	AAQ41001 Human plasma carbo
8	889	46.5	1400	21	AAF18005 Lung cancer associ
9	640.5	33.5	1215	16	AAQ90600 Porcine Tyr-His-Me
10	639	33.4	1263	17	AAT42494 Human pancreatic c
11	639	33.4	1263	19	AAV41795 Human pancreatic c
12	631.5	33.0	1284	17	AAT42506 ProHCPB gene with
13	623	32.6	1254	18	AAV17331 PreproHCPB-linker-
14	621.5	32.5	1332	24	ABL65831 Lung cancer relate
15	618.5	32.4	2023	25	ABZ75116 Anti-human seminal
16	611.5	32.0	1302	25	ABV77254 Nucleotide sequenc
17	611.5	32.0	1311	22	AAO38337 Human carboxypepti
18	611.5	32.0	1993	24	AAO30579 Human protease, PR
19	604	31.6	927	17	AAT35760 Rat mature carboxy
20	591	30.9	416	25	ABX49430 Bovine EST associa
21	576	30.1	921	16	AAQ90601 Porcine carboxypep
22	564	29.5	2128	22	AAO38339 Human carboxypepti
23	563.5	29.5	1378	25	ABT33355 NOVX DNA sequence
24	563	29.5	1125	22	AAH76477 Human secreted met
25	563	29.5	1125	22	AAH76479 cDNA encoding huma
26	563	29.5	1125	24	ABK31745 DNA encoding novel
27	563	29.5	1222	25	ABT33353 NOVX DNA sequence
28	563	29.5	1342	22	AAH76476 Human secreted met
29	563	29.5	1342	22	AAH76478 Human secreted met
30	563	29.5	1344	25	ABQ77393 Human zinc-binding
31	563	29.5	1603	24	AAO39053 Human 23566 (carbo
32	563	29.5	1826	22	AAF81728 Human protease and
33	561	29.4	999	17	AAT42500 Mature HCPB coding
34	561	29.4	1053	17	AAT42497 mature HCPB-(His)6
35	554	29.0	1053	20	AAZ4804 Human carboxypepta
36	554	29.0	1059	17	AAT42511 Modified HCPB (D25
37	554	29.0	1059	18	AAT62787 Carboxypeptidase B
38	553	28.9	1059	17	AAT42512 Modified HCPB (D25
39	553	28.9	1059	18	AAT62788 Carboxypeptidase B
40	550	28.8	1059	18	AAT62806 Carboxypeptidase B
41	548	28.7	1059	18	AAT62790 Carboxypeptidase B
42	548	28.7	1059	18	AAT62801 Carboxypeptidase B
43	547	28.6	1059	18	AAT62789 Carboxypeptidase B
44	547	28.6	1196	25	ABT33352 NOVX DNA sequence
45	546	28.6	1059	18	AAT62791 Carboxypeptidase B

ALIGNMENTS

RESULT 1

AAC81962
ID AAC81962 standard; cDNA; 1573 BP.

AC AAC81962;

XX
DT 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B cDNA.

DE Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimer; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.

XX Homo sapiens.
OS

XX Key Location/Qualifiers
 FH 18..1100
 FT /*tag= a
 FT /product= "carboxypeptidase B"
 XX WO200066717-A1.
 XX 09-NOV-2000.
 XX 01-MAY-2000; 2000WO-JP02878.
 XX 30-APR-1999; 99JP-0125169.
 XX (MATS/) MATSUMOTO A.
 XX Matsumoto A;
 XX WPI; 2000-687534/67.
 DR P-PSDB; AAB11457.
 XX Human brain carboxypeptidase B isolated from the hippocampus useful for
 FT screening agents for the treatment of Alzheimer's and other brain
 PT disorders -
 XX Claim 2b; Page 64-68; 84pp; Japanese.
 CC This invention describes a novel protein with peptidase activity
 CC against brain beta-amyloid precursor protein which has been isolated from
 CC human hippocampus and which has cerebroprotective, antialzheimers,
 CC neurotropic, neuroprotective and hemostatic activity and which can be used
 CC as a vaccine or for gene therapy. The protein, and compounds identified
 CC by screening as promoters or inhibitors of its activity, are used to
 CC regulate beta-amyloid accumulation in the brain and treat or prevent
 CC diseases in which this occurs, such as Alzheimer's, senile dementia,
 CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
 XX Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;
 SQ

Alignment Scores:
 Pred. No.: 7.2e-195 Length: 1573
 Score: 1911.00 Matches: 360
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-980-881A-2 (1-360) x AAC81962 (1-1573)

Qy	1	MetLysLeuCysSerLeuAlaValLeuValProLleValLeuPheCysGluGlnHisVal	20
Db	18	ATGAAGCTTTTCAGACCTTGCGAGTCTTGATGCCATGTTCTCTCTGAGCAGCATGTC	77
Qy	21	PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln	40
Db	78	TTTCGGTTCCAGAGTGCCAGTTCTAGTCTCTCTCTAGAACCTCTAGGCAAGTTCAA	137
Qy	41	ValLeuGlnAnLeuThrThrThrTyrGluileValLeuTrpGlnProValThrAlaAsp	60
Db	138	GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGCGACCGGTAACAGCTGAC	197
Qy	61	LeuileVallyLysGlnValHisPhePheValAlaLeuSerAspValAspAnVal	80
Db	198	CTTATTGTGAAGAAAACAAGTCAATTTTTTTGTAATGTCATCTGATGTCGACAAATGTG	257
Qy	81	LysAlaHisLeuAnValSerGlyLeuProCysSerValLeuLeuAlaAspValcLuAsp	100
Db	258	AAAGCCATTAAATGTAGCGGAATCCATGCGAGTCTTGCTGGCAGAGCTGGAAGAT	317
Qy	101	LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr	120
Db	318	CTTATTCAACAGCAGATTTTCAACGACACAGTACGTCGCCGCCGAGCTCCGCACTGCTACTAT	377

Qy	121	GluGlnTyrHisSerLeuAnGluileTyrSerTrpIleGluPheIleThrGluArgHis	140
Db	378	GAACAGTATCACTCACTAAATGAAATCTATCTTGGATAGATTTTATAACTGAGAGCAT	437
Qy	141	ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal	160
Db	438	CCTGATATGCTTACAAAAATCCACATTTGGATCTCTCATTTGAGAAAGTACCCACTCTATGTT	497
Qy	161	LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle	180
Db	498	TTAAAGGTTTCTGAAAAAGAACCAACAGCAAAATGCCATATGATTTGATCTGGAATC	557
Qy	181	HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet	200
Db	558	CATGCCAGAGATGGATCTCTCTGCTTTTCTGTTGTGTTTCATAGGCCATAATCGAATG	617
Qy	201	TrpArgLysAnArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAnSer	220
Db	618	TGGAGAAAGAACCGTCTCTTCTATCGCAACATCATTTGCATCGGAACAGACCTGAATAGC	677
Qy	221	AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr	240
Db	678	AACTTTCTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCCTCATGCTCGAAACC	737
Qy	241	TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg	260
Db	738	TACTGTGACTTTATCTCTGAGTCAGAACCCAGAGTGAAGCAGTGGTGTCTTCTTGAGA	797
Qy	261	ArgAsnIleAnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal	280
Db	798	AGAAATATCAACAGATTAAGCATACATCAGCATGCATTCATCATCTCCAGCATATAGTG	857
Qy	281	PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValala	300
Db	858	TTTCCATATTTCTATACACGAAGTAAAGCAAGACCATCAGGAACTGTCTCTAGTAGCC	917
Qy	301	SerGluAlaValArgAlaIleAspLysThrSerLysAnThrArgTyrThrHisGlyHis	320
Db	918	AGTGAGCAGTTCTGTCTATTGACAAAATAGTAAAAATACCAGGTATACACATGGCCAT	977
Qy	321	GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGly	340
Db	978	GGCTCAGAAACCTTATACCTAGCTCTCTGAGGTGGGACGATTGGAATCTATGATTGGGC	1037
Qy	341	IleLysTyrSerPheThrSerAnProValGlnLysLeuLeuProLeuSerLeuLys	360
Db	1038	ATCAATATTTCTGTTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAA	1097

RESULT 2
 ABX13670
 ID ABX13670 standard; cDNA; 1625 BP.
 XX
 AC ABX13670;
 XX
 DT 14-FEB-2003 (first entry)
 XX
 DE Human protease cDNA.
 XX
 KW Human; gene; ss; protease; proteolytic degradation; proteolysis;
 KW proliferation; differentiation; signalling; therapeutic; gene therapy;
 KW protein therapy; diagnostic; immune response; vaccine; inflammation;
 KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
 KW single nucleotide polymorphism; SNP.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..16
 FT /*tag= a
 FT CDS 17..1177
 FT /*tag= b
 FT /product= "Protease"
 FT variation replace(521,A)

XX AAV74302;
 AC
 XX
 DT 28-APR-1999 (first entry)
 XX
 DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
 XX
 KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
 KW polymorphism detection; thrombotic disease; ds.
 XX
 OS Homo sapiens.
 XX
 FN WO985645-A1.
 XX
 PD 10-DEC-1998.
 XX
 PF 02-JUN-1998; 98WO-EP03244.
 XX
 PR 03-JUN-1997; 97US-0869057.
 XX
 PA (SCHD) SCHERING AG.
 XX
 PI Morser MJ, Nagashima M;
 XX
 DR WPI; 1999-045800/04.
 DR
 DR P-PSDB; AAW92270.
 XX
 XX Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 PT humans
 XX
 PS Example 1; Page 24; 35pp; English.
 XX
 CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
 CC hPCPBthr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphs of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.
 XX
 SQ Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;
 XX

Alignment Scores:
 Pred. No.: 1272
 Score: 1832.00 Length: 1272
 Percent Similarity: 86.27% Matches: 357
 Best Local Similarity: 86.02% Mismatches: 2
 Query Match: 95.87% Indels: 55
 DB: 20 Gaps: 2

US-09-980-881A-2 (1-360) x AAV74302 (1-1272)

Oy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
 Db 1 ATGAAGCTTTGCAGCTTGCAGTCCCTGTGACCATGTTCTCTCTGAGCAGCATGTC 60
 Oy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
 Db 61 TTGCGGTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTCTAGAACTCTAGCAAGTTCAA 120
 Oy 41 ValLeuGlnAsnLeuThrThrThrTyxGluIleValLeuThrProValThrAlaAsp 60
 Db 121 GTTCTACAGAATCTTACTACACATATGAGATTGTTCTCTGCGACCGCGTAAACAGCTGAC 180
 Oy 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAsnVal 80

Db CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTGCAATGTG 240
 Oy LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
 Db AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGGCAGACGTGGAAGAT 300
 Oy LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyxTyx 120
 Db CTTATTCAACAGCAGCATTTCCACGACACAGTCAGCCCCGAGCCTCCGCAATCGTACTAT 360
 Oy GluGlnTyxHisSerLeuAsnGluIleTyxSerTrpIleGluPheIleThrGluArgHis 140
 Db GACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAACTGAGAGCAT 420
 Oy ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyxProLeuTyxVal 160
 Db CCTGATATGCTTACAAAAATCCACATTTGGATCCCTCATTTGAGAAAGTACCCCACTCTATGTT 480
 Oy LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
 Db TTAAGGTTTCTGGAAGAAGAACAAACAGCCAAAATGCCATATGCGATTGACTGTGGAATC 540
 Oy HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
 Db CATGCCAGAGAAATGGAATCTCTCTGCTTCTCTGCTTGTGTTTCATAGGCCATATAAACA 600
 Oy ----- 197
 Db TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660
 Oy -----AsnArgMetTrpArgLys 203
 Db CCGGTGTTAATGTGGACGGTTATGACTACTATGGAAGAAAGATCGAATGTGGAGAAAG 720
 Oy AsnArgSerPheTyxAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
 Db AACCGTCTTCTATGCGAACATCATTTGCATCGAACAGACCTGAATAGGAATTTGCT 780
 Oy SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThyTyxCysGly 243
 Db TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840
 Oy LeuTyxProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
 Db CTTTATCTGAGTCAGAACACAGAAAGTGAAGCAGTGGCTAGTTCTTGTGAGAAAGAAATATC 900
 Oy AsnGlnIleLysAlaTyxIleSerMetHisSerTyxSerGlnHisIleValPheProTyx 283
 Db AACCAGATTAAAGCATACATCAGCATTCATCTCCAGCATATAGTGTTCATAT 960
 Oy SerTyxThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
 Db TCCTATACAGAGTAAAGCAAAAGACCATGAGGAACCTGTCTCTAGTAGCAGGTGAAGCA 1020
 Oy ValArgAlaIleAspLysThrSerLysAsnThrArgTyxThrHisGlyHisGlySerGlu 323
 Db GTTCGTGCTATTGAGAAACTAGTAAATAATACCAAGGTATACATGCGCATGGCTCAGAA 1080
 Oy ThrLeuTyxLeuAlaProGlyGlyGlyAspAspTrpIleTyxAspLeuGlyIleLysTyx 343
 Db ACCTTATACCTAGTCTCTGGAGGTGGGACGATGGATCTATGATTTGGGCATCAAAATAT 1140
 Oy SerPhe-----Th 346
 Db TCGTTTCAATTGAATTCGAGATACGGGCACATACGGAATTTCTTGTCTGCGGAGCGTTAC 1200
 Oy xSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
 Db ATCAAAACCCACTGTAGAGAGCTTTTGGCGCTGTCTCTATAAA 1243
 Oy

RESULT 4
 ABN95899

RESULT 5

AA111671

ID AA111671 standard; DNA; 1749 BP.

XX AC

AA111671;

XX XX

25-MAR-2003 (updated)

DT 12-APR-1996 (first entry)

XX XX

DE Human plasma carboxypeptidase B coding sequence.

XX KW

KW Plasma carboxypeptidase B; hPCPB; antibody; detection;

KW purification; plasminogen; affinity column; ss.

XX XX

OS Homo sapiens.

XX FH

Key Location/Qualifiers

CDS 41..1312

FT /*tag= a

FT /product= Human plasma carboxypeptidase B.

FT /*tag= b

FT mat_peptide 107..1309

FT /*tag= c

XX PN

US5474901-A.

XX XX

12-DEC-1995.

XX PD

19-JUL-1994; 94US-0277540.

XX XX

01-FEB-1991; 91US-0649591.

XX PR

14-OCT-1992; 92US-0959944.

XX PR

15-DEC-1993; 93US-0167727.

XX PR

19-JUL-1994; 94US-0277540.

XX XX

(GETH) GENENTECH INC.

XX PA

Drayna DT, Eaton DL;

XX PI

WPI; 1996-039508/04.

XX DR

P-PSDB; AAR90293.

XX XX

PT Antibody to human plasma carboxypeptidase B - useful for detecting

PT and purifying hPCPB for use in treating clotting disorders e.g.

XX PT

haemophilia A

XX PS

Disclosure; Figure 4; 40pp; English.

XX PS

CC An antibody which specifically binds human plasma carboxypeptidase B (hPCPB) and does not cross react with other carboxypeptidases is

CC useful for the detection of hPCPB in vitro. The antibody is also

CC used for purifying hPCPB from a sample. Purification comprises

CC passing a sample thought to contain hPCPB over either a column to

CC which antibody has been bound, or a plasminogen affinity column,

CC eluting the column and then recovering the fraction containing the

XX CC

CC hPCPB.

XX CC

CC (Updated on 25-MAR-2003 to correct PF field.)

XX SQ

Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

XX SQ

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Db	41	ATGAAGCTTTGACGCTTGGAGTCTCTTGTATCCCATTTCTCTTCTGTGACGAGTATGTC	100
Qy	21	PheAlaPheGlnSerGlyGlnValValLeuAlaLeuProArgThrSerArgGlnValGln	40
Db	101	TTTCGGTTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA	160
Qy	41	ValLeuGlnAenLeuThrThrThrTyrGluLeuValLeuTyrGlnProValThrAlaAsp	60
Db	161	GTTCACAGATCTTACTACCAATATGAGATTGTTCTTGGCAGCCGTTACAGCTGAC	220
Qy	61	LeuileVallyLysLysGlnValHisPheValAsnAlaSerAspValAspVal	80
Db	221	CTTATTGTGAAGAAAAACCAAGTCCATTTTGTAAATGCATCTGATGTGCAAAATGTG	280
Qy	81	LysAlaHisLeuAenValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAsp	100
Db	281	AAAGCCCATTTAAATGTGAGCGGAATTCATGCGAGTGTCTTGTGGCAGAGCTGGAAGAT	340
Qy	101	LeuileGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr	120
Db	341	CTTATTCAACAGCAGATTTCCACGACACAGTCAGCCCCGAGCTCCGCACTGCTACTAT	400
Qy	121	GluGlnTyrHisSerLeuAenGluIleTyrSerTyrIleGluPheIleThrGluArgHis	140
Db	401	GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGATTATATACTGAGAGCAT	460
Qy	141	ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal	160
Db	461	CCTGATATGCTTACAAAAATCCCATTTGGATCCCTCATTTGAGAAAGTACCCACTCTATGTT	520
Qy	161	LeuLysValSerGlyLysGluGlnThrAlaLysAenAlaIleTyrIleAspCysGlyIle	180
Db	521	TTAAAGGTTTCTGAAAGAAACAAACAGACGATGCAATATGCAATATGCAATATGCAATATG	580
Qy	181	HisAlaArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis	197
Db	581	CATGCCAGAGATGATCTCTCTGCTTCTGCTTGTGTTTCATAGGCCATATACTCAA	640
Qy	197	-----	197
Db	641	TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG	700
Qy	198	-----	198
Db	701	CCGTTGTTAATGTGGAGGTTATGACTACTATCGGAAAGAAATCGAATGTGGAGAAAG	760
Qy	204	AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAenSerAsnPheVal	223
Db	761	AACCGTTCTTCTATGCGAACAAATCATTTGCATCGAACACAGCTGAATAGGAACCTTGCT	820
Qy	224	SerLysHisTyrCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly	243
Db	821	TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGA	880
Qy	244	LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle	263
Db	881	CTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTGTGAGAGAAATATC	940
Qy	264	AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr	283
Db	941	AACCAAGATTAAAGCATATACATCAGCATGCTATCTCCAGCATATAGTGTTCATAT	1000
Qy	284	SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla	303
Db	1001	TCCTATACAGCAAGTAAAGCAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA	1060
Qy	304	ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu	323
Db	1061	GTTCGTGCTATTGAGAAACTAGTAAATAACACAGGTATACATGCGCATGGCTCGAA	1120
Qy	324	ThrLeuTyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyr	343

US-09-980-881A-2 (1-360) x AA111671 (1-1749)

Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

```

Db 1121 ACCTTATACCTAGCTCCTGGAGTGGGACGATGGATCTATGATTTGGGCATCAATAT 1180
Qy 344 SerPhe-----Th 346
Db 1181 TCGTTTAAATTAAGCTTCGAGATACGGGCACATACGATTTCTTGCTCGCGGAGGTTTAC 1240
Qy 346 rSerAsnProValGluLeuLeuProLeuSerLeuLys 360
Db 1241 ATCAAAACCCACTGTAGAGAGAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 6
ID AAT62846 standard; DNA; 1749 BP.
AC AAT62846;
XX
DT 25-MAR-2003 (updated)
DT 08-MAY-1997 (first entry)
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW plasma; plasminogen; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 41..1312
FT /tag= a
FT /product= Human PCPB
FT sig_peptide 41..106
FT mat_peptide 107..1309
FT misc_binding 134..177
FT /tag= d
FT /bound_moeity= 46_bp_probe

XX US5593674-A.
XX
XX 14-JAN-1997.
XX
XX 27-APR-1995; 95US-0430787.
XX
XX 01-FEB-1991; 91US-0649591.
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX 27-APR-1995; 95US-0430787.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX WPI; 1997-099413/09.
XX P-ESDB; AAW14733.
XX
XX Using human plasma carboxypeptidase B in blood coagulation - is
XX functionally related to carboxypeptidase A and pancreas
XX carboxypeptidase B
XX
XX Example 2; Column 37-42; 39pp; English.
XX
XX This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
XX PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX purified from human plasma or by transformed cell culture by
XX extraction using plasminogen bound to a solid phase.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
XX
Alignment Scores:

```

```

Pred. No.: 2.54e-186 Length: 1749
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 18 Gaps: 2

US-09-980-881A-2 (1-360) x AAT62846 (1-1749)

Qy 1 MetLysLeuCySerSerLeuAlaValLeuValProIleValLeuPheCyGluGlnHisVal 20
Db 41 ATGAAGCTTTTGACGCTTGCAGTCTCTGTATCCCATTTGTTCTCTCTGTGACGACATGTC 100
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 101 TTCGCGTTTCAGAGTGCGCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 160
Qy 41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
Db 161 GTTCTACAGAACTTTACTACAACATATGAGATTCTTCTGGCAGCCGGTAACAGCTGAC 220
Qy 61 LeuIleValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 80
Db 221 CTTATTGTGAAGAAAAACAGTCCATTTTGTAAATGCAATCTGATGTGACAAATGTG 280
Qy 81 LysAlaHisLeuAsnValSerGlyIleProCySerValLeuLeuAlaAspValGluAsp 100
Db 281 AAAGCCCATTTAAATGTGAGCGGAATTCATGCACTGCTTGTGCGCAGACGTTGGAAGAT 340
Qy 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 341 CTTATTCAACAGCAGATTTCCACGACACAGTACGCCCGCAGCGCTCGTACTAT 400
Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
Db 401 GAACAGTATCTACTCACTAAATGAATCTATTCTTGGATAGAAATTTTAACTGAGAGGAT 460
Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 461 CCTGATATGCTTACAAAAATCCACATTTGGATTCCTCATTTGAGAAAGTAGTCCCACTCTATGTT 520
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Db 521 TTTAAAGGTTTCTGAAAAAGAACAAACAGCAGCAATGCCATATGGATTTGATGTGGAATC 580
Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 581 CATGCCAGAGAAATGGATCTCTCTGCTTTCTGCTGTGGTTTCATAGGCCATATAACTCAA 640
Qy 197 ----- 197
Db 641 TTCTATGGGATATAGGGCAATATACCAATCTCTCAGGCTTGTGGATTTCTATGTTATG 700
Qy 198 -----AsnArgMetTrpArgLys 203
Db 701 CCGGTGGTTAATGTGACGGTTTATGACTACTCATCGAAAAAGAAATCGAAATGTGGAAG 760
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 761 AACCGTTCTTTCTATGCGAACAAATCATTTGCATCGGAACAGACCTCGAATAGGAACCTTGC 820
Qy 224 SerLysHisTrpCyGluGluGlyValAspSerSerSerCySerGluThrTyrCysGly 243
Db 821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCGGAACCTACTGTGA 880
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
Db 881 CTTTATCTCGAGTCAGAACACAGAGAGTGAAGCAGTGGCTAGTTCTTTTCAGAGAAGAAATC 940
Qy 264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 941 AACCAAGATTAAAGCATACATCAGCATGTCATCTCATCTCCAGCATATAGTGTTCATAT 1000

```

Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
 Db 1001 TCCTATACAGAGTAAAGCAAGACCATGAGGAACCTGTCTAGTAGCCAGTGAAGCA 1060
 Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
 Db 1061 GTTCGTGCTATTGAGAAACTAGTAAAAATACCAAGGTATACACATGGCCATGGCTCAGAA 1120
 Qy 324 ThrLeuTyrLeuAlaProGlyGlyClyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
 Db 1121 ACCTTATACCTAGCTCTCTGGAGGTGGGACGATTGATCTATGATTTGGGCATCAAAATAT 1180
 Qy 344 SerPhe-----Th 346
 Db 1181 TCGTTTACAATTGAATTCGAGATACGGGCACATACCGATTCTTGCTGCCGGAGGTTTAC 1240
 Qy 346 rSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
 Db 1241 ATCAAAACCCAGCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAAA 1283

RESULT 7

ID AAQ41001 standard; cDNA; 1749 BP.

AC AAQ41001;

DT 24-AUG-1993 (first entry)

XX Human plasma carboxypeptidase B gene.

XX PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT sig_peptide 41..106

FT mat_peptide 107..1312

FT misc_feature 134..177

FT /product= PCPB

FT /tag= c

FT /function= probe

FT /note= "used to obtain full-length clones"

XX US206161-A.

XX 27-APR-1993.

XX 01-FEB-1991; 91US-0649591.

XX 01-FEB-1991; 91US-0649591.

XX (GETH) GENENTECH INC.

XX Drayna DT, Eaton DL;

XX WPI; 1993-151724/18.

XX P-PSDB; AAR36273.

XX New human plasma carboxypeptidase B - used as haemostatic
 PT regulator for clotting blood, partic. for treating blood clotting
 PT disorders, e.g. haemophilia
 XX Disclosure; Fig 4; 40pp; English.

XX Human plasma carboxypeptidase B was isolated from human plasma and
 CC partially sequenced. Oligonucleotide primers were designed based on
 CC the partial amino acid sequences. The primers were used in a PCR
 CC amplification to identify cDNA encoding PCPB from a human liver cDNA
 CC library. The PCR product was capable of encoding the first 37 amino
 CC acids of PCPB; a 46mer probe was used to obtain the full-length
 CC sequence which, although disclosed in the specification, is not

CC claimed. PCPB inhibits the enzymatic conversion by tPA of
 CC plasminogen to plasmin in the presence of fibrinogen.

XX Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

Alignment Scores:

Pred. No.: 8.73e-186 Length: 1749
 Score: 1827.00 Matches: 356
 Percent Similarity: 86.02% Conservative: 1
 Best Local Similarity: 85.78% Mismatches: 3
 Query Match: 95.60% Indels: 55
 DB: 14 Gaps: 2

US-09-980-881A-2 (1-360) x AAQ41001 (1-1749)

Qy 1 MetLysLeuCyssSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
 Db 41 ATGAAGCTTTGCGAGCTTGCACTCTTGTATGCCATTTCTTCTCTGTGAGCAGCATGTC 100
 Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
 Db 101 TTCGCGTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGCAAGTTCAA 160
 Qy 41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
 Db 161 GTTCTACAGAATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220
 Qy 61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
 Db 221 CTTATTGTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTGATGTGCAATGTG 280
 Qy 81 LysAlaHisLeuAsnValSerGlyIleProCyssSerValLeuLeuAlaAspValGluAsp 100
 Db 281 AAAGCCCATTTAAATGTGAGCGGAATTCATGACAGTCTTGTGCGAGACGTGGAAGAT 340
 Qy 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
 Db 341 CTTATTCAACAGCAGATTTCCACAGCACACAGTCAGCCCCCGAGCTCGCATCTGACTAT 400
 Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
 Db 401 GAACAGTATCACTCACTAAATGAAATCTATTCTTTGGATAGAAATTTATAAATGAGAGCAT 460
 Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
 Db 461 CCTGATATGCTTCAAAAAATCCACATTTGATGATTCCTATTGAGAAAGTACCCACTCTATGTT 520
 Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
 Db 521 TTAAGGTTTCTGGAAAGAACAAACAGCCAAAATGCCATATGGATTTGACTGTGGAAATC 580
 Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
 Db 581 CATGCCAGAGAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGGCCATATAAATCAA 640
 Qy 197 ----- 197
 Db 641 TTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 700
 Qy 198 -----AenArgMetTrpArgLys 203
 Db 701 CCGGTGTTAATGTGGACGGTTATGACTACTCATCGAAAAAAGAAATCGAATGTGGAAAG 760
 Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
 Db 761 AACCGTTCTTTCTATGCGAAACAATCATTTGCATCGGAACAGACACCTGAATAGAACTTGTCT 820
 Qy 224 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 243
 Db 821 TCCAAACACCTGGTGTGAGGAAGGTGCATGTCAGTTCCTCATGCTCGGAACCTACTGTGGA 880
 Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263

Db 881 CTTTATCTGAGTCAGAACAGAGTGAAGCAGTGGCTAGTTCTTTCAGAGAAATATC 940
Qy 264 AenGlnIleLysAlaTyrlleSerMetHisSerTyrlleSerGlnHisIleValPheProTyrlle 283
Db 941 AACAGATTAAAGCATACATCAGCATGATTCATCTCCAGCATATAGTGTTCATAT 1000
Qy 284 SerTyrlleSerSerLysSerLysSerLysSerLysSerLysSerLysSerLysSerLysSerLys 303
Db 1001 TCCATATACAGAAAGTAAAGCAAGAACCATGAGGAACCTGTCTAGTAGCCAGTGAAGCA 1060
Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrlleSerGlnHisGlySerGlu 323
Db 1061 GTTCGTGTATGAGAAACTAGTAAATACCATGATATACATGCGCTGCATGAGCA 1120
Qy 324 ThrLeuTyrlleAlaProGlyGlyGlyAspPheTyrlleTyrlleTyrlleTyrlleTyrlle 343
Db 1121 ACCTTATACCTAGCTCTCTGGAGTGGGACGATGATGATGATGATGATGATGATGATGATGAT 1180
Qy 344 SerPhe-----Th 346
Db 1181 TCGTTTACAACTTTCGAGATACGGGCACATACGGAATCTTCTCCCGGAGCGTTAC 1240
Qy 346 rSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1241 ATCAACCCACTGTAGAGAGCTTTTCCGCTGTCTCTAANA 1283

RESULT 8
AAFI18005
ID AAFI18005 standard; DNA; 1400 BP.
XX AAFI18005;
AC AAFI18005;
XX
XX 14-MAR-2001 (first entry)
XX Lung cancer associated polynucleotide sequence SEQ ID 24.
XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX Homo sapiens.
XX
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
XX P-PSDB; AAB58129.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX Claim 1; Page 507; 1425pp; English.
XX
XX Polynucleotide sequences AAFI17982 - AAFI18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the

CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAFI18425 - AAFI18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
Alignment Scores:
Pred. No.: 3e-85 Length: 1400
Score: 889.00 Matches: 178
Percent Similarity: 75.21% Conservative: 1
Best Local Similarity: 74.79% Mismatches: 3
Query Match: 46.52% Indels: 56
DB: 21 Gaps: 2
US-09-980-881A-2 (1-360) x AAFI18005 (1-1400)
Qy 179 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--- 197
Db 10 GGAATCCATGCCAGANAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 69
Qy 197 ----- 197
Db 70 ACTCAATTCATGGGATAAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTAT 129
Qy 198 -----AsnArgMetTr 201
Db 130 GTTATGCCGGTGTAAATGCGATGTTATGAACTACTCATGGAAGAAAGATCGAATGTG 189
Qy 201 pArgGlyAsnArgSerPheTyrlleAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAs 221
Db 190 GAGAAGAAGACCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 249
Qy 221 nPheValSerLysHisTrpCysGluGluValSerSerSerSerSerSerSerSerSerSerSer 241
Db 250 CTTTCTTCCAAACACCTGTGTGAGGAGGTGCATCCAGTTCCTCATCTCGGAACCTA 309
Qy 241 rCysGlyLeuTyrlleProGluSerGluProGluValIleAlaValAlaSerPheLeuArgAr 261
Db 310 CTGTGGACTTATCTCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTCTGAGAG 369
Qy 261 gAsnIleAsnGlnIleLysAlaTyrlleSerMetHisSerTyrlleSerGlnHisIleValPh 281
Db 370 AATATCAACCAAGATTAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGT 429
Qy 281 eProTyrlleSerTyrlleArgSerLysSerLysAspHisGluLysLeuSerLeuValAlaSe 301
Db 430 TCCATATTCCTATACAGAGTAAAGCAAGCAAGTGAAGCAAGTGTCTCTAGTAGCCAG 489
Qy 301 rGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrlleHisGlyHisG 321
Db 490 TGAAGCAGTCTGCTGCTATTTGAGAAACTAGTAAATAATACAGGATATACATGGCCATGG 549
Qy 321 ySerGluThrLeuTyrlleAlaProGlyGlyGlyAspAspTrpIleTyrlleTyrlleTyrlle 341
Db 550 CTCAGAAACCTTATACCTAGCTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCAT 609
Qy 341 eLysTyrlleSerPhe----- 345
Db 610 CAATATTCGTTTACAATTTGAATTCGAGATACGGGCACATACGGATTTCTTGTCTCCGGA 669
Qy 346 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
Db 670 GCGTTACATCAACCCACTGTAGAGAGCTTTTCCGCTGTCTCTAANA 719

RESULT 9

AAQ90600
 ID AAQ90600 standard; DNA; 1215 BP.
 XX
 AC AAQ90600;
 XX
 13-MAR-1996 (first entry)
 XX
 DT
 XX
 DE Porcine Tyr-His-Met procarboxypeptidase B coding sequence.
 XX
 KW Procarboxypeptidase B; carboxypeptidase B; pichia; PCPB;
 KW human serum albumin; premating factor alpha; mating factor alpha;
 KW proCBB; ds.
 XX
 OS Sus scrofa.
 XX
 FH Location/Qualifiers
 FT 1..1215
 FT /*tag= a
 FT /product= Porcine procarboxypeptidase B.
 XX
 PN W09514096-A1.
 XX
 PD 26-MAY-1995.
 XX
 PF 16-NOV-1994; 94WO-US13142.
 XX
 PR 16-NOV-1993; 93US-0153258.
 XX
 PA (ELIL) LILLY & CO BLI.
 XX
 PI Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
 PI Sterner JL, Zhang H;
 XX
 DR WPI; 1995-200386/26.
 DR P-PSDB; AAR75131.
 XX
 PT DNA encoding porcine carboxypeptidase B - used for transforming
 PT host cells, partic. Pichia species, for prodn. of the enzyme
 XX
 PS Claim 2; Page 20-21; 34pp; English.
 XX
 CC The porcine carboxypeptidase B coding sequence can be place in a
 CC bacterial or pref. Pichia yeast expression vector. The expression
 CC vector further comprises the signal peptide of either human serum
 CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
 CC (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
 CC pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
 CC (designated pLGD27 - NRRL B-21027). The method can be used for
 CC producing large amounts of porcine carboxypeptidase B and when
 CC produced in Pichia yeast, the protein does not need solubilisation
 CC or folding. The produced enzyme is then used for pref. cleaving
 CC basic residues from the carboxy terminus of proteins.
 XX
 SQ Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;
 XX

Alignment Scores:
 Pred. No.: 1,14e-58 Length: 1215
 Score: 640.50 Matches: 136
 Percent Similarity: 54.29% Conservative: 60
 Best Local Similarity: 37.67% Mismatches: 122
 Query Match: 33.52% Indels: 43
 DB: 16 Gaps: 6

US-09-980-881A-2 (1-360) x AAQ90600 (1-1215)

Qy 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
 |||||
 Db 28 TTCGAAGGGGAGAGGTGTTCCGTCTCAATGTTGAAGATGAATGACATCAGCTTACT 87
 |||||
 Qy 43 GlnAsnLeuThrThrTyrrGluLeuValLeuThrGlnProValThrAlaAspLeu 62
 |||||
 Db 88 CATGAGTTGGCCAGCACCAGGAGGATGCTCTGGAACACAGATTCTGTGCACACAAATC 147
 |||||

Qy 63 ValLysLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 82
 |||||
 Db 148 AACCTCACAGTACAGTTGACTTCCGTGTGAAGACAGATATTTTGGCTGTGGAAGAC 207
 |||||
 Qy 83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 102
 |||||
 Db 208 TTTCTGGAGCAGATGAATACATACATATGAGTACTCATAAACACCTGAGATCTGTGCTC 267
 |||||
 Qy 103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrrGluGln 122
 |||||
 Db 268 GAGCTCAGTTTGCACAGAGATC-----CGTAACTGGACACAGTTATGAGAAG 318
 |||||
 Qy 123 TyrHisSerLeuAsnGluIleTyrrSerTrpIleGluPheIleThrGluArgHisProAsp 142
 |||||
 Db 319 TACAACAACCTGGGAACGATCGAGGCTTGGACTAAGCAAGTCACCACTGAAATCCAGAC 378
 |||||
 Qy 143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrrProLeuTyrrValLeuLys 162
 |||||
 Db 379 CTCATCTCTGCACAGCCATCGGAACACTACATTTTAGGAAACAATATATATACCTCTCAAG 438
 |||||
 Qy 163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
 |||||
 Db 439 GTT---GGCAAACTGGACCAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 495
 |||||
 Qy 183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 195
 |||||
 Db 496 AGAATGGATTTCCCATGCTATTTTGGCAGTGGTTTGTGAGAGAGGCTGTCTCACCTAT 555
 |||||
 Qy 195 ----- 195
 |||||
 Db 556 GGATATGAGAGTCACATGACAGAAATTCCTCAACAGCTAGACTTTTATGTCTTCCCTGTG 615
 |||||
 Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 205
 |||||
 Db 616 CTCATATATGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGAGAAAGACCCGC 675
 |||||
 Qy 206 SerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
 |||||
 Db 676 TCTACCAATGCTGGAACTACTGCTGATGGACAGACCCCAACAGAAATTTT--GATGCT 732
 |||||
 Qy 226 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrrCysGlyLeuTyrr 245
 |||||
 Db 733 GGGTGGTGCACTGAGGAGCTCTACAGACCCCTGCCATGAGACTTACTGTGGATCTGCT 792
 |||||
 Qy 246 ProGluSerGluProGluValLysAlaValAspSerPheLeuArgArgAsnIleAsnGln 265
 |||||
 Db 793 GCAGAGTCTGAAAAGAGACCAAGGCCCTGGCTGATTTATACGCAACACCTCTCTCTCC 852
 |||||
 Qy 266 IleLysAlaTyrrIleSerMetHisSerTyrrSerGlnHisIleValPheProTyrrSerTyrr 285
 |||||
 Db 853 ATCAAGCATACCTGAGTCCACTCATCTACAGATGATATCTTACCTTATTCCTAT 912
 |||||
 Qy 286 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 305
 |||||
 Db 913 GATTACAAACTCCCGAGAACAAATGCTGAGTTGAATAACCTGGCTAAGGCTGCGTGAAA 972
 |||||
 Qy 306 AlaIleAspLysThrSerLysAsnThrArgTyrrThrHisGlyHisGlySerGluThrLeu 325
 |||||
 Db 973 GAACCT---GCTACACTGTATGGCACCAGTACATACATACGAGGAGGCTTACACAAATC 1029
 |||||
 Qy 326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrrAspLeuGlyIleLysTyrrSerPhe 345
 |||||
 Db 1030 TATCTGCTGCTGGGGGCTCTGATGACTGGCTTATGACCAAGGAAATCAATATATCTCTTC 1089
 |||||

Qy 346 Thr 346
 |||||
 Db 1090 ACC 1092

RESULT 10
 AAT42494
 ID AAT42494 standard; DNA; 1263 BP.
 XX
 AC AAT42494;

Db 1066 GCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATCTCTCAC 1122

RESULT 11

AAV41795
ID AAV41795 standard; DNA; 1263 BP.

AAV41795:

20-NOV-1998 (first entry)

DE Human pancreatic carboxypeptidase nucleotide sequence.

ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW prodruq therapy.

OS Homo sapiens.

XX	Key	Location/Qualifiers
FH		

FT	key	LOCATION/
FT	CDS	1..1248

```
FT
/*tag= a
```

FT /product= "Pancreatic carboxypeptidase B"

FT /note= "No start codon given"

PN WO9835988-A1.

PD 20-AUG-1998.

10-FEB-1998; 98WO-GB00415.

PR 29-OCT-1997: 97GB-0022727

PR 14-FEB-1997: 97GB-0003104

PR 18-OCT-1997; 97GB-0022003;

PA (ZENE) ZENEPA (ZENE) ZENECA LTD.

PI Edge MD:

WPI: 1998-467168/40.

DR P-PSDB; AAW74476.

PT New modified pro-domain of carboxy-peptidase B - enhances expression
PT of co-expressed proteins for production of recombinant
PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT enzyme prodrgug therapy

PS Example 1; Page 52-53; 83pp; English.

The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a modified pro-domain of CPB on a separate gene to enhance recombinant expression. This process can be used to produce recombinant CPB in eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in insulin production and protein sequencing, while its fusions with antibodies are useful in antibody-directed enzyme prodrug therapy. The modified pro-domain provide increased yields of recombinant CPB, possibly by protecting the C-terminus against enzymatic degradation or by increasing intracellular trafficking.

SQ Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;

Alignment Scores:

Pred. No.:	1,766-58	Length:	1263
Score:	639.00	Matches:	135
Percent Similarity:	53.8%	Conservative:	69
Best Local Similarity:	35.6%	Mismatches:	131
Query Match:	33.4%	Indels:	44
DB:	19	Gaps:	7

US-09-980-881A-2 (1-360) x AAV41795 (1-1263)

Qy 6 LeuAlaValLeuValProIleValLeuPheCysGluHisValPheAla--PheGln 24

D_b 4 CTC TTG GTT CTG GTG ACT GTG GGC CCT GGC ATC TGC TCAT CAT GGT GTG AGC ACT TTG AA 63

RESULT 12

AAT42506
ID AAT42506 standard; DNA; 1284 BP.

AC AAT42506;

DT 17-FEB-1997 (first entry)

XX PROHCPB gene with PelB leader sequence.

XX ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; Ffab'; PelB; leader; human carboxypeptidase B; ss.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..1275

FT /tag= a

FT sig_peptide 1..66

FT /tag= b

FT sig_peptide /note= "PelB leader sequence"

FT /tag= c

FT /note= "pro sequence"

FT mat_peptide 352..1272

FT /tag= d

FT /note= "mature HCPB sequence"

XX W09620011-A1.

XX 04-JUL-1996.

XX 21-DEC-1995; 95WO-GB02991.

XX 16-AUG-1995; 95GB-0016810.

XX 23-DEC-1994; 94GB-0026192.

XX (ZENE) ZENECA LTD.

XX Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW, AM;

XX Hennem JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;

XX Tarragona-Fiol A, Taylorson CJ;

XX WPI; 1996-321650/32.

XX P-PSDB; AAW06175.

XX Two component system for anti-tumour therapy - comprising targeting

XX moiety linked to mutated enzyme which can transform an

XX anti-neoplastic prodrug

XX Reference Example 18; Page 140-142; 182pp; English.

XX A two-component system for anti-tumour therapy comprises a targeting
XX moiety linked to a mutated enzyme which can transform an anti-neoplastic
XX prodrug. The system is based on antibody directed enzyme prodrug therapy
XX (ADEPT) using non-naturally occurring mutant forms of host enzymes,
XX pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
XX Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
XX be used. The present sequence is a cloned pro-HCPB gene contained in
XX plasmid pIC11738 and which can be expressed in E. coli.

XX Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;

Alignment Scores:

Score: 1.15e-57 Length: 1284
Percent Similarity: 631.50 Matches: 129
Best Local Similarity: 54.57% Conservative: 68
Query Match: 35.73% Mismatches: 121
Indels: 43
Gaps: 17

US-09-980-881A-2 (1-360) x AAT42506 (1-1284)

Qy 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
Db 85 TTTGAAGCGGAGAAGGTGTTCCGTGTTAACTTGAAGATGAAATCAATTAACATAATC 144

Qy 43 GlnAsnLeuThrThrThrGluLeuValLeuTrpGlnProValThrAlaAspLeu 62
Db 145 CGCGAGTTGGCCAGCAGCAGCAGATTCTCTGGAAGCCAGATTCTGTGCACACAAATC 204

Qy 63 VallyslyslsGlnValHisPheValAsnAlaSerAspValAspValAspValysala 82
Db 205 AAACCTCACAGTACAGTTGACTTCGTGTTAAAGCAGAAAGATCTGTCTGTGGAGAAAT 264

Qy 83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 102
Db 265 GTTCTAAAGCAGAAATGAATACATACAGGTACTGATAAGCAACCTGAGAAATGTGGTG 324

Qy 103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
Db 325 GAGGCTCAGTTTGATGACCGGTT-----CGTGCACAGCAGCAGATTATGAGAAG 375

Qy 123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
Db 376 TACAACAAGTGGGAAACGATAGAGCTTGGGACTCAACAAGTCGCCACTGAGAAATCCAGCC 435

Qy 143 MetLeuThrIleHisIleGlySerSerPheGluIleTyrProLeuTyrValLeuLys 162
Db 436 CTCATCTCTGCAGTGTATTCGGAACCAACATTTGAGGGACCGCTATTACTCTCTGAAG 495

Qy 163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisala 182
Db 496 GTT---GGCAAGCTGGACAAAATAAGCCTGCCATTTTCATGGAGCTGGTTTCCATGCC 552

Qy 183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 195
Db 553 AGAGAGTGGATTTCTCTCGCATTTGCCAGTGGTTTGTAAAGAGAGGCTGTTCGTACCTPAT 612

Qy 195 ----- 195
Db 613 GGACGTGAGATCCAAAGTGACAGAGCTTCTGCACAAAGTTAGACTTTTATGTCTCGCTGTG 672

Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 205
Db 673 CTCAATATTGATGCTACATCTACACCTGGACCAAGAGCCGATTTTGAGAAAGACTCGC 732

Qy 206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
Db 733 TCCACCCATACTGGATCTAGCTGCTGATGGCACAGACCCCAACAGAAATTTT---GATGCT 789

Qy 226 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 245
Db 790 GGTGGTGTGAATTTGGAGCCTCTCGAAACCCCTGTGATGAACCTTACTGTGGACCTGCC 849

Qy 246 ProGluSerGluProGluValLysAlaValAspSerPheLeuArgArgAsnIleAsnGln 265
Db 850 GCAGAGTCTGAAAGGAGAGACCAAGSCCTGCTGATTTTCATCCGCAACAACTCTCTTCC 909

Qy 266 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 285
Db 910 ATCAAGGCATATCTGACAATCCATCCGCTACTGCCAAATGATGATGATCTACCTTACTCAT 969

Qy 286 ThrArgSerLysSerLysAspHisGluLuleuSerLeuValAlaSerGluAlaValArg 305
Db 970 GCTTACAACTCGGTGAGAACCAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1029

Qy 306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 325
Db 1030 GAACCTT---GCCTCACTGCACGGCACCACCAAGTACATATATGCCCCGGAGCTACAAATC 1086

Qy 326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345
Db 326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345

RESULT 14

ABL65831
ID ABL65831 standard; DNA; 1332 BP.
XX AC
XX ABL65831;
XX
DT 15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:4168.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytosaratic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-233133P.
XX 18-SEP-2000; 2000US-233617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234059P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237606P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.

XX

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set -

Claim 1; SEQ ID 4168; 44pp; English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's tumour.

XX SQ Sequence 1332 BP; 357 A; 321 C; 311 G; 343 T; 0 other;

Alignment Scores:

Pred. No.:	1.43e-56	Length:	1332
Score:	621.50	Matches:	134
Percent Similarity:	53.66%	Conservative:	71
Best Local Similarity:	35.08%	Mismatches:	132
Query Match:	32.52%	Indels:	45
DB:	24	Gaps:	8

US-09-980-881A-2 (1-360) x ABL65831 (1-1332)

Qy	3	LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla	22
Db	24	ATGTTGGCACTCTTGGTTCTGGTGACTGGCCCTGGCATCTGCTCATCTGTTGGTGG	83
Qy	23	---PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnVal	41
Db	84	CACCTTTGAAGCGAGAAGGTTCCTCGTTTAAAGATGAATAATCATCAATTAACATA	143
Qy	42	LeuGlnAsnLeuThrThrThrGluIleValLeuTrpGlnProValThrAlaAspLeu	61
Db	144	ATCCGCGAGTTGGCCACACGACCAGATTGACTTCTGGAAGCCAGATTCTGTCAACAA	203
Qy	62	IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys	81
Db	204	ATCAAACTCACAGTACAGTTGACTTCGGTGTAAAGCAGAAGATCTGTCACGTGGAG	263
Qy	82	AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu	101
Db	264	AATGTTCTAAAGCAGAAGTAACTACAATAACAAGGTACTGTATAGCAACCTGAGAAATGTG	323
Qy	102	IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrrGlu	121
Db	324	GTGAGGCTCAGTTTGTATAGCCGGGTT-----CGTCAACAGACAGATTTATAG	374
Qy	122	GlnTyrrHisSerLeuAsnGluIleTyrrSerTrpIleGluPheIleThrGluArgHisPro	141
Db	375	AAGTACAACAAGTGGGAACGATAGAGCTTGGACTCAACAAGTCGCCCATGAGATCCA	434
Qy	142	AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrrProLeuTyrrValLeu	161
Db	435	GCCTCATCTCTCGCAGTGTATTCGGAACCAACATTTGAGGACGCGGTATTATCTCTCTG	494
Qy	162	LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyLeHis	181
Db	495	AAGGTT---GGCAAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGTTCCAT	551

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1849.5	96.8	1625	4	US-09-813-133A-1	Sequence 1, Appli
2	1832	95.9	1272	4	US-08-869-057-1	Sequence 1, Appli
3	1832	95.9	1749	1	US-07-849-591B-2	Sequence 2, Appli
4	1832	95.9	1749	1	US-08-377-540-2	Sequence 2, Appli
5	1832	95.9	1749	1	US-08-430-787A-2	Sequence 2, Appli
6	640.5	33.5	1215	1	US-08-696-139-1	Sequence 1, Appli
7	639	33.4	1263	4	US-08-860-882A-56	Sequence 56, Appl
8	639	33.4	1263	4	US-09-011-769A-38	Sequence 38, Appl
9	631.5	33.0	1284	2	US-08-860-882A-71	Sequence 71, Appl
10	631.5	33.0	1284	4	US-09-011-769A-55	Sequence 55, Appl
11	623	32.6	2154	3	US-09-171-945-124	Sequence 124, App
12	611.5	32.0	1311	4	US-09-675-305-9	Sequence 9, Appli

Db 121 GTTCTACAGATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
Qy 61 LeulleVallyLysGlnValHiePheValAenAlaSerAspValAspVal 80
Db 181 CTTATTGTGAAGAAAAAACAAGTCCATTTTGTAAATGTCATCTGATGTCGACAAATGTG 240
Qy 81 LysAlaHisLeuAenValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAsp 100
Db 241 AAAGCCATTAAATGTGAGGGAAATCCATGCGAGTGTCTTGTGGCAGAGCTGGAAGAT 300
Qy 101 LeulleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 301 CTTATTCAACAGCAGATTTCACACACACAGTCAGCCCCGAGGCTCCGATCTGACTAT 360
Qy 121 GluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHis 140
Db 361 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAAATGAGAGCAT 420
Qy 141 ProAspMetLeuThrLysIleHisGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 421 CCGTATATGCTTACAAAATCCATGATGATCCCTCATTTGAGAAGTACCCCACTCTATGTT 480
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Db 481 TTAAGGTTTCTGAAAAGAACAAACACGCCAAAATGCCATATGATGATGACTGGAAATC 540
Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 541 CATGCCAGAGATGGATCTCTCTCTGCTTCTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT 600
Qy 197 ----- 197
Db 601 TTCTATGGGATATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 660
Qy 198 -----AsnArgMetTrpArgLys 203
Db 661 CCGTGTGTTAATGTGGAGGTTATGACTACTCATCGAAAAGAAATCGAATGTGGAAAG 720
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 721 AACGGTCTTCTATGCGAACCAATCATTTGTCATCGAAGACAGACCTGAATAGGAATTTGCT 780
Qy 224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
Db 781 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCATGCTCGGAAACCTACTGTGA 840
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
Db 841 CTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTGGAGAAGAAATATC 900
Qy 264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 901 AACAGATTAAAGCATACATCAGCATGCAATTCATCTCCAGCATATAGTGTTCATAT 960
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 961 TCCTATACCAAGTAAAGCAAGCAAGCAATGAGCAACTGTCTCTAGTAGCCAGTGAAGCA 1020
Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1021 GTTCGTGCTATTGAGAAACATAGTAAATAATACAGGTATATACATGCGCATGGCTCAGAA 1080
Qy 324 ThrLeuTyrLeuAlaProGlyGlyCysAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
Db 1081 ACCTTATACCTAGCTCTCTGAGGAGTGGGACCAATGGATCTATGATTTGGGCATCAATAT 1140
Qy 344 SerPhe-----Th 346
Db 1141 TCGTTTAACTTGAATTCGAGATACGGGCACATACGGATTCCTGCTCCCGGAGCGTTAC 1200
Qy 346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1201 ATCAAAACCCACCTGTAGAGAGCTTTTGGCCGCTGTCTCTAAAA 1243

RESULT 3

US-07-649-591B-2
: Sequence 2, Application US/07649591B
: Patent No. 5206161
: GENERAL INFORMATION:
: APPLICANT: Dennis Drayna and Daniel Eaton
: TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/649,591B
: FILING DATE: 19910201
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hasek, Janet E.
: REGISTRATION NUMBER: 28,616
: REFERENCE/DOCKET NUMBER: 689
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/266-1896
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1749 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: hybridization probe
: LOCATION: 133 to 178
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: FEATURE:
: NAME/KEY: potential clip site
: LOCATION: 380 to 382
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
: FEATURE:
: NAME/KEY: signal sequence
: LOCATION: 41 to 106
: IDENTIFICATION METHOD:
: OTHER INFORMATION:
US-07-649-591B-2
Alignment Scores: 3.58e-219 Length: 1749
Pred. No.: 1832.00 Matches: 357
Score: 86.27% Conservative: 1
Percent Similarity: 86.02% Mismatches: 2
Best Local Similarity: 95.87% Indels: 55
Query Match: 1 Gaps: 2
DB:
US-09-980-881A-2 (1-360) x US-07-649-591B-2 (1-1749)

Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 41 ATGAAGCTTTGACGCCCTTGACGCTCTGTGCCCATTTGTTCTTCTGTGAGCAGCATGTC 100

Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 101 TTGGGTTTCAGAGTGGCCAAAGTTCTAGCTCTCTCTAGAACCTCTAGCAAGTTCAA 160
Qy 41 ValLeuGlnAenLeuThrThrTyrTyrGluLeuValLeuTrpGlnProValThrAlaAsp 60
Db 161 GTTCTACAGAATCTTACTACACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220
Qy 61 LeuLeuValLeuLysGlnValHisPhePheValAenAlaSerAspValAspVal 80
Db 221 CTTATTGTTGAAGAAAAACAAGTCCATTTTTTTGTAATGTCATCTGATGTCGACAAATGTG 280
Qy 81 LysAlaHisLeuAenValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAsp 100
Db 281 AAAGCCCATTTAATGTGAGGGGAATTCATGCGAGTGTCTTGTGGCAGAGCTGGAAGAT 340
Qy 101 LeuLeuGlnGlnLeuSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 341 CTTATTCAACAGCAGATTTCACACGACACAGTCCAGCCCGAGCCTCCGCACTCGTACTAT 400
Qy 121 GluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpLeuGluPheLeuThrGluArgHis 140
Db 401 GAAAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAAAGTGGAGGCAT 460
Qy 141 ProAspMetLeuThrLysLeuHisLeuGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 461 CTTGATATGCTTACAAAATCCACATTTGGATCTCTCAATTTGGAAGTACCCACTCTATGTT 520
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaLeuTrpLeuAspCysGlyLeu 180
Db 521 TTAAGGTTTCTGGAAAGAACAAACAGCCAAAATGCCATATGGAATGAGTGTGGAATC 580
Qy 181 HisLysGluTyrTrpLeuSerProAlaPheCysLeuTrpPheLeuGlyHis 197
Db 581 CATGCCAGAAATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATATAACTCAA 640
Qy 197 ----- 197
Db 641 TTCTATGGGAATATAGGCCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 700
Qy 198 ----- AsnArgMetTrpArgLys 203
Db 701 CCGGTGGTTAATGTGACGGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGAGAAAG 760
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysLeuGlyThrAspLeuAenSerAsnPheVal 223
Db 761 AACCGTTCTTTCTATGCGAACCAATCATTTGCATCGGAACAGACACCTGGAATAGGAATTTGCT 820
Qy 224 SerLysHisTrpCysGluGluGluValaSerSerSerCysSerGluThrTyrCysGly 243
Db 821 TCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCTCATGCTCGGAACCTACTGTGGA 880
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValaSerPheLeuArgArgAsnLeu 263
Db 881 CTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTTTGGAGAAGAAATATC 940
Qy 264 AsnGlnLeuLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 941 AACCAAGATTAAGCATATACATCAGCATGCAATTCATCACTCCAGCATATAGTGTTCATAT 1000
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 1001 TCCATACACAGATTAAGCAAGACCATGAGGACTGTCTCTAGTAGCCAGTGAAGCA 1060
Qy 304 ValArgAlaLeuAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1061 GTTCGTGCTATTGAGAAACTAGTAAAAATACCAGGTATACATACATGCGCATGTGCTCAGAA 1120
Qy 324 ThrLeuTyrLeuAlaProGlyGlyGlyAspSerTrpLeuTyrAspLeuGlyLeuLysTyr 343
Db 1121 ACCTTATACCTAGCTCTCTGGAGGTGGGACCAATGGATCTATGATTTTGGGCATCAATAT 1180
Qy 344 SerPhe-----Th 346

Db 1181 TCGTTTACAATTGAACCTTCGAGATACGGGCACATACGGAATTTCTTGCTGCCGACGCTTAC 1240
Qy 346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1241 ATCAAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTATAAA 1283
RESULT 4
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-277-540-2
Alignment Scores:
Pred. No.: 3.58e-219 Length: 1749
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 1 Gaps: 2
US-09-980-881A-2 (1-360) x US-08-277-540-2 (1-1749)
Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 41 ATGAAGCTTTGACGCCCTTGACCTTGATCCCATTTGTTCTCTCTGTGACACATGTC 100
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 101 TTGCGTTTTCAGAGTGGCCAAAGTTCTAGCTCTCTCTAGAACCTCTAGCAAGTTCAA 160

Qy 41 ValLeuGlnAenLeuThrThrThyTyrGluLeuValLeuTrpGlnProValThrAlaAsp 60
Db 161 GTTCTACAGAACTTACTACAACTATGAGATTGTTCTTGCGCAGCGGTAAACAGCTGAC 220
Qy 61 LeuLeuValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 80
Db 221 CTTATTGTGAAGAAAACAAGTCACTTTTGTAAATGATCTGATGTCGACATGTG 280
Qy 81 LysAlaHisLeuAenValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAsp 100
Db 281 AAAGCCATTAAATGTGAGCGGAATCCATGAGTGTCTTGCTGGCAGAGCTGGAAGAT 340
Qy 101 LeuLeuGlnGlnHisSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 341 CTTATTCAACAGCAGATTTCCAAAGCACACAGTACGCCCCGAGCCTCCGCATCGTACTAT 400
Qy 121 GluGlnTyrHisSerLeuAenGluLeuTyrSerTrpLeuGluPheLeuThrGluArgHis 140
Db 401 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAACTGAGAGCAT 460
Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 461 CCTGATATGCTTCAAAAAATCCACATTTGGATCTCTCAITTTGAGAGTACCCCACTATGTT 520
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaLeuTrpLeuAspCysGlyIle 180
Db 521 TTAAGGTTTCTGAAAGAACAAACAGCCAAAAATGCCATATGCAATGAGTGTGAATC 580
Qy 181 HisAlaArgGluTyrPheSerProAlaPheCysLeuTrpPheLeuGlyHis----- 197
Db 581 CATGCCAGAGAATGGATCTCTCTGCTTCTGCTTGTGCTTGTGTTTCATAGGCCATATAACTCAA 640
Qy 197 ----- 197
Db 641 TTCTATGGATAATAGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATG 700
Qy 198 -----AsnArgMetTrpArgLys 203
Db 701 CCGGTGGTTAATGTGACGGTTATGACTACTCATGGAAAAAGAAATCGAATGTGGAGAAAG 760
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAenSerAsnPheVal 223
Db 761 AACGGTCTTCTATGCGAAACAATCATTTGCATCGCAACAGACCTGAATAGCAACTTTGCT 820
Qy 224 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 243
Db 821 TCCAAACACTGGTGTGAGGAGGTGCATCCAGTTCTCTGCTCGGAACCTACTGTGGA 880
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIle 263
Db 881 CTTTATCTGAGTCAGAACACAGAGAGTGAAGCAGTGGCTAGTTCTTTGAGAGAAATATC 940
Qy 264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 941 AACAGATTAAGCATACATCAGATGCTTCACTATCTCCAGCATATAGTGTTCCTCATAT 1000
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAla 303
Db 1001 TCCATACAGAAATAAAGCAAGACATGAGAACTGCTCTAGTAGCCAGTGAAGCA 1060
Qy 304 ValArgAlaLeuAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAAGGTATACATGCGCCATGGCTCGAA 1120
Qy 324 ThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
Db 1121 ACCTTATACCTAGCTCCTGGAGGTGGGACGATTTGGATCTATGATTTTGGGCATCAAAATAT 1180
Qy 344 SerPhe-----Th 346
Db 1181 TCGTTTACAATTGAATTCGAGATACGGGCAATACCGGATTTCTTGTGCGCGGACGGTTAC 1240

Qy 346 rSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1241 ATCAAAACCCCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAAA 1283
RESULT 5
US-08-430-787A-2
; Sequence 2, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-430-787A-2
Alignment Scores:
Pred. No.: 3-58e-219 Length: 1749
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 1 Gaps: 2

US-09-980-881A-2 (1-360) x US-08-430-787A-2 (1-1749)
Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 41 ATGAAGCTTTGCAGCCTTGAGTCTTGTATCCCATTTCTTCTCTGAGCAGCATGTC 100
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 101 TTCGCGTTTCAGATGGCCAAAGTTCTAGCTGCTTCTCTAGAACCTCTTAGGCAAGTTCAA 160

Qy 63 VallysLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 82
Db 148 AAACCTCACAGTACAGTTCGGTGTGAAGCAGAGATATATTTGGCTGTGGAAGAC 207
Qy 83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLeu 102
Db 208 TTTCTGGAGCAGATGAATCAATATCAGGTATCTATAACACACCTGAGATCTGTGCTC 267
Qy 103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyTyTyGluGln 122
Db 268 GAGGCTCAGTTTGACAGCAGATC-----CGTACAACCTGGACACAGTTATGAGAAG 318
Qy 123 TyrHisSerLeuAsnGluIleTySerTrpIleGluPheIleThrGluArgHisProAsp 142
Db 319 TACAACAACTGGGAAACGATCGAGCTTGGACTAAGCAAGTACACAGTGAATAATCCAGAC 378
Qy 143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyTyProLeuTyValLeuLys 162
Db 379 CTCATCTCTCCACAGCCATCGGAATCACTATTTTAGGAACAATATATACCTCTCTCAG 438
Qy 163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
Db 439 GTT---GGCAAACTGGACCAATAAAGCTGCCATTTTCATGGAGTGTGGTTTCCATGCC 495
Qy 183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 195
Db 496 AGAGAATGGATTTCCCATGCACTTTTGGCAGTGGTTTGTGAGAGAGGCTGTCTCACCTAT 555
Qy 195 ----- 195
Db 556 GGATATGAGAGTCAATGACAGAAATTCCTCAACAGCTAGACTTTTATGTCTTCCCTGTG 615
Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 205
Db 616 CTCAATATTGATGGCTACATCTACACCTGGACCAAGACCGAATGTGAGAGAACCCGCG 675
Qy 206 SerPheTyAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
Db 676 TCTACCAATGTGGAACCTACCTGCAATGGCACAGACCCCAACAGAAATTTT---GATGCT 732
Qy 226 HisTrpCysGluGluGlyValSerSerSerCysSerGluThrTyTyCysGlyLeuTy 245
Db 733 GGGTGGTGCAACACTGGAGGCTCTACAGACCCCTGGATGAGACTTACTGTGGATCTGCT 792
Qy 246 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 265
Db 793 GCAGAGTCTGAAGAAGAGACCAAGCCCTGCTGATTTTATACGCAACACCTCTCTCC 852
Qy 266 IleLysAlaTyIleSerMetHisSerTySerGlnHisIleValPheProTyTySerTy 285
Db 853 ATCAAGCATACCTGACGATCCACTCATCTACTACTACAGATGATCTACTACCTTATTCCTAT 912
Qy 286 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 305
Db 913 GATTACAACCTCCCGAGAAACAATGCTGAGTTGAATAACCTGGTAAAGCTGCGGTGAAA 972
Qy 306 AlaIleAspLysThrSerLysAsnThrArgTyTyThrHisGlyHisGlySerGluThrLeu 325
Db 973 GAACCT---GCTACACTATGGCACCAGTACATACAGCCGCCAGGAGCTACACCAATC 1029
Qy 326 TyrLeuAlaProGlyGlyAspAspTrpIleTyAspLeuGlyIleLysTySerPhe 345
Db 1030 TATCTGCTGTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAATAATTCCTTC 1089
Qy 346 Thr 346
Db 1090 ACC 1092

RESULT 7

US-08-860-882A-56

; Sequence 56, Application US/08860882A

; Patent No. 5985281

; GENERAL INFORMATION:

APPLICANT: TAYLORSON, CHRISTOPHER JOHN
APPLICANT: EGGLETE, HENDRIKUS JOHANNES
APPLICANT: TARRAGONA-PIOL, ANTONIO
APPLICANT: RABIN, BRIAN ROBERT
APPLICANT: BOYLE, FRANCIS THOMAS
APPLICANT: HENNAM, JOHN FREDERICK
APPLICANT: BLAKELY, DAVID CHARLES
APPLICANT: MARSHAM, PETER ROBERT
APPLICANT: HEATON, DAVID WILLIAM
APPLICANT: DAVIES, DAVID HUW
TITLE OF INVENTION: CHEMICAL COMPOUNDS
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 1263 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-56

Alignment Scores:
Pred. No.: 6,34e-70 Length: 1263
Score: 639.00 Matches: 135
Percent Similarity: 53.83% Conservative: 69
Best Local Similarity: 35.62% Mismatches: 131
Query Match: 33.44% Indels: 44
DB: 2 Gaps: 7

US-09-980-881A-2 (1-360) x US-08-860-882A-56 (1-1263)

Qy 6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla---PheGln 24
Db 4 CTCTTGTGTCTTGGTGTGCTGTGGCCCTGCTCATCATGTGTGTGTGACACTTTGAA 63
Qy 25 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsn 44
Db 64 GCGCAGAGAGGTGTTCCTGTGTTAAGATGAAAATCACATTAAACATTAATCCGCGAG 123
Qy 45 LeuThrThrThrTyTyGluIleValLeuTrpGlnProValThrAlaAspLeuIleValLys 64
Db 124 TTGGCCAGCAGCAGCCAGATTGACTTCTTGAAGCCAGATTCTGTGCACACAATCAACCT 183
Qy 65 LysLysGlnValHisPhePheValAsnAlaSerValAspAsnValLysAlaHisLeu 84
Db 184 CACAGTACAGTTGACTTCCCGTGTAAAGCAGAGATATCTGTCACTGTGGAGAAATGTTCTA 243
Qy 85 AsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 104
Db 244 AAGCAGAATGAACCTACAACTACAGGTACTGATAGCAACCTGAGAAATGTGTGGAGGCT 303

Db 415 TCTCGAGTGTATCGGAACACATTTTGGGAGCGGCTATTACCTCCTCGAAGTT--- 471
Qy 165 GlyLysGluGlnThrAlaLysAsnAlaLeuTrpLeuAspCysGlyLeuHisAlaArgGlu 184
Db 472 GGCAGAGCTGACAAAATAAGCTGCCATTTTCATGGACTGTGGTTTCCATGCCAGAGAG 531
Qy 185 TrpLeuSerProAlaPheCysLeuTrpPheLeu----- 195
Db 532 TGGATTTCTCTGCAATTCGCAAGTGGTTTGTAAAGAGAGGCTGTGTGACCTATGGAGCT 591
Qy 195 ----- 195
Db 592 GAGATCCAAAGTGACAGAGCTTCTGCAAGTTAGACTTTTATGTCTCCTGCTGCTCAAT 651
Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsnArgSerPhe 207
Db 652 ATTGATGCTCATCATACACCTGGACCAAGAGCGGATTTTGGAGAAAGACTCGCTCCACC 711
Qy 208 TyrAlaAsnAsnHisCysLeuGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
Db 712 CATACTGGATCTAGCTGCATTTGGCACAGACCCCAACAGAAATTTT---GATGCTGGTGG 768
Qy 228 CysGluGluGlyAlaSerSerCysSerGluThrTrpCysGlyLeuTrpProGlu 247
Db 769 TGTGAATTGGAGCTCTCGAAACCCCTGTGATGAATTTACTGTGGACCTGCCGACAG 828
Qy 248 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnLeuAsnGlnLeuLys 267
Db 829 TCTGAAAGGAAACCAAGGCCCTGCTGATTTTCATCGCAACAAACTCTCTTCATCAAG 888
Qy 268 AlaTrpLeuSerMetHisSerTrpSerGlnHisLeuValPheProTrpSerTrpArg 287
Db 889 GCATATCTGCAAAATCCACTCTGCTACTCCCAATGATGATCTACCCCTTACTCATATGCTTAC 948
Qy 288 SerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArgAlaLeu 307
Db 949 AAATCGGTGAGACAAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGGAAGAAGTT 1008
Qy 308 AspLysThrSerLysAsnThrArgTrpThrHisGlyHisGlySerGluThrLeuTrpLeu 327
Db 1009 ---GCCTCACTGCACGGCACCAAGTACATATGCGCGGAGCTACAACTATCTATCCT 1065
Qy 328 AlaProGlyGlyLysAspTrpLeuTrpLeuGlyLysGlyLysSerPheThr 346
Db 1066 GCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTCCACC 1122

RESULT 9

US-08-860-882A-71
; Sequence 71, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 1284 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-71
Alignment Scores:
Pred. No.: 5,67e-69 Length: 1284
Score: 631.50 Matches: 129
Percent Similarity: 54.57% Conservative: 68
Best Local Similarity: 35.73% Mismatches: 121
Query Match: 33.05% Indels: 43
DB: 2 Gaps: 6
US-09-980-881A-2 (1-360) x US-08-860-882A-71 (1-1284)
Qy 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
Db 85 TTTGAAGCGCAGAGAGGTTCCTCGTTTAACTGGAAGTGAATCAATTAACATAATC 144
Qy 43 GlnAsnLeuThrThrTrpGluLeuValLeuTrpGlnProValThrAlaAspLeuLeu 62
Db 145 CGCGAGTGGCCAGCAGCAGCCAGATTGACTTCTGGAAGCCAGATTCTGTCCACACAATC 204
Qy 63 ValLysLysGlnValHisPheValAsnAlaSerAspValAsnValLysAla 82
Db 205 AAACCTCAGATCAGTTGACTTCCGTGTTAAAGCAGAAATATCTGCTGAGGAAT 264
Qy 83 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeuLeu 102
Db 265 GTTCTAAAGCAGATGAATACATACAGTACTGATAGCAACCTGAGAAATGTGTG 324
Qy 103 GlnGlnGlnSerAsnAspThrValSerProArgAlaSerAlaSerTrpTrpGluGln 122
Db 325 GAGGCTCAGTTTGTATAGCCGGTT-----CGTGCACACAGGACACAGTTATGAGAAG 375
Qy 123 TyrHisSerLeuAsnGluLeuTrpSerTrpLeuPheLeuThrGluArgHisProAsp 142
Db 376 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTGGCCACTGAGAAATCCAGCC 435
Qy 143 MetLeuThrLysLeuHisLeuGlySerSerPheGlyLysTrpProLeuTrpValLeuLys 162
Db 436 CTCATCTCTCGCAGTGTATATCGGAACACATTTGAGGGACCGCGCTATTACTCTCTGAAG 495
Qy 163 ValSerGlyLysGluGlnThrAlaLysAsnAlaLeuTrpLeuAspCysGlyLeuHisAla 182
Db 496 GTT---GCCAAAGCTGGACAAAATAAGCTGCCATTTTTCATGGACTGTGGTTTCCATGCC 552
Qy 183 ArgGluTrpLeuSerProAlaPheCysLeuTrpPheLeu----- 195
Db 553 AGAGAGTGGATTTCTCTGCAATTCGCAAGTGGTTTGTAAAGAGAGGCTGTTCGTACCTAT 612
Qy 195 ----- 195
Db 613 GGAGCTGAGATCCAAAGTGACAGAGCTTCTCGCAAGTTAGACTTTTATGTCTGCTGCTGTG 672
Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 205

```
Db 673 CTCATATTGATGGCTACATCTACACCTGGACCAAGCCGATTTTGGAGAAAGACTGCG 732
Qy 206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnAsnValSerLys 225
Db 733 TCCACCCACTAGTGGATCTAGCTGCTATGGCAGACCCCAACAGAAATTTT---GATGCT 789
Qy 226 HistpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 245
Db 790 GGTGGTGTAATGGAGCCCTCTCGAAACCCCTGTGATGAACCTTACTGTGGACCTGCC 849
Qy 246 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 265
Db 850 GCAGAGTCTGAAGAGACCAAGCCCTGCTGATTTTCATCCGCAACAACTCTCTTCC 909
Qy 266 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 285
Db 910 ATCAAGGCATATCTGACAACTCTGCTACTCTCCCAATGATGATCTACCTTACTCATAT 969
Qy 286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 305
Db 970 GCTTACAACTCGGTGAGAACATCTGAGTTGAAATGCCCTGGCTAAAGCTACTGTGAAA 1029
Qy 306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 325
Db 1030 GAACCT---GCCTCACTGCACGGCACCAAGTACACATATGCGCCGGGAGCTACAACTC 1086
Qy 326 TyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345
Db 1087 TATCTGCTGCTGGGGCTCTGACAGCTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1146
Qy 346 Thr 346
Db 1147 ACC 1149

RESULT 10
US-09-011-769A-55
; Sequence 55, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; CORRESPONDENCE ADDRESS:
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; ADDRESS: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011,769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
```

```
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1284 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1272
; NAME/KEY: mat peptide
; LOCATION: 352..1272
; SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-011-769A-55

Alignment Scores:
Pred. No.: 5,67e-69 Length: 1284
Score: 631.50 Matches: 129
Percent Similarity: 54.57% Conservative: 68
Best Local Similarity: 35.73% Mismatches: 121
Query Match: 33.05% Indels: 43
DB: 4 Gaps: 6

US-09-980-881A-2 (1-360) x US-09-011-769A-55 (1-1284)
Qy 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
Db 85 TTTGAAGCGGAGAGGTGTTCCTCGTTAAAGTTGAAGTGAATAATCATTAACTAATC 144
Qy 43 GlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuLeu 62
Db 145 CGCAGTTGGCCAGCAGCACCCAGATTGACTTCTGGAAGCCAGATTCTGTGCACAAATC 204
Qy 63 ValLysLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 82
Db 205 AAACCTCACAGTACAGTTGACTTCCGTTTAAAGCAGAGATATCTGTCTGTGGAGAT 264
Qy 83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLeu 102
Db 265 GTTCTAAAGCAGATGAATCACTACATAACAAGTACTGATAAGCAACCTGAGAAATGTGTG 324
Qy 103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
Db 325 GAGCTCAGTTTGTATAGCCGGTT-----CGTCAACAGGACACAGTTTATGAGAAG 375
Qy 123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
Db 376 TACAACAAAGTGGGAAACGATAGAGCTTGGACTCAACAAGTCCGCTGAGATCCAGCC 435
Qy 143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
Db 436 CTCATCTCTCGCAGTGTATCGGAACCAACATTTGAGGGACGCGCTATTATCTCTCTGAAG 495
Qy 163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
Db 496 GTT---GGCAAGCTGGACAAATAAGCTGCCATTTTCATGGACTGTGGTTTCCATGCC 552
Qy 183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 195
Db 553 AGAGAGTGGATTTCTCTCGCATCTGCGAGTGGTGTGTAAGAGAGGCTGTTCGTACCTAT 612
Qy 195 ----- 195
Db 613 GGACGTGAGATCCAAAGTCAAGAGCTTCTCGCAAGTTAGACTTTTATGTCTCGCTGTG 672
Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 205
Db 673 CTCAATATTGATGGCTACATCTACCTGGACCAAGAGCCGATTTTGGAGAAAGCTGCG 732
Qy 206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
```


SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/782.760

FILING DATE: 13-JAN-1997

CLASSIFICATION: 435

CLASSIFICATION: 433
PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: 08/378.233

APPLICATION NUMBER: 08/3
FILING DATE: 25-JAN-1995

FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION

NAME: White, John P.
REGISTRATION NUMBER: 28 678

REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0226/42842

REFERENCE/DOCKET NUMBER: 03-
TELECOMMUNICATION INFORMATION

TELECOMMUNICATION INFORMATION
TELEPHONE (212) 338-3400

TELEPHONE: (212) 278-0401

TELEFAX: (212) 391-0525

FORMATION FOR SEQ ID NO: 5

SEQUENCE CHARACTERISTICS:

LENGTH: 927 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

NAME/REF: LOCATION:

LOCATION.
-782-760-5

C-00/-78/-

ment Scores:

ment scores:	No.:	Length:	927
9 110-66			

No.:	9,11e-66	927
Length:		125
Matches:	504.00	

Matches:	125
604.00	
Concomitant:	44
60.148	

Local Similarity:	60.14%	Conservative:	44
Local Similarity:	44.40%	Misclassification:	70

US-09-980-881A-2 (1-360) x US-08-782-760-5 (1-927)

115	AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAenGluLleTyrSerTrpIleGlu	134
Db		
1	GCAAGTGGACACAGCTACACCAAGTACAAACATGGGAAACGATTGAGGCGTGCAATCAA	60
Db		
135	PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu	154
Db		
61	CAAGTTGCCACTGATAATCCAGACCTTGCTACTCAGAGCGTCATTGGAACCACTTTGAA	120
Db		
155	LysTyrProLeuTyrValLeuLysValSerGlyLysGluInThrAlaLysAsnAlaIle	174
Db		
121	GGAGCTGAACATGTATGTCCTCAAGATT--GGTAAACTAGACCGAATAAGCGTCGCATC	177
Db		
175	TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe	194
Db		
178	TTCATCGATTGTGGTTTCATGSCAAGAGAGTGSATTTCTCTGCATTCTGTCAGTGGTTT	237
Db		
195	Ile-----	195
Qy	:::	
238	GTGAGAGGCGTCCGTACCTTAATCAAGAGATCCACATGAACAGCTTCTAGATGAA	297
Db		
196	-----GlyHis-----	197
Qy		
298	CTGGATTTCATGTTCTGCCTGTGCTCAACATTGATGGCTATGTCTACACCTGGACTAAG	357
Db		
198	AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp	217
Qy		
358	GACAGAAATGTGAGAAAAACCGCTCTACTATGCTTGAAGTTCCTCTTGGGTGTAGAC	417
Db		
218	LeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaLaserSerSerCys	237
Qy		
418	CCCAACAGGAATTTT---AATGCTGGTGTGTGAATGGGAGCTTCTCGGAGTCCCTGC	474
Db		
238	SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer	257
Qy		

Score: 604.00 Matches: 125
Percent Similarity: 60.14% Conservative: 44
Best Local Similarity: 44.48% Mismatches: 72
Query Match: 31.61% Indels: 41
DB: 5 Gaps: 5

US-09-980-881A-2 (1-360) x PCT-US96-00995-5 (1-927)

Qy 115 AlaSerAlaSerTyrTyrGluGlnTyrHisLeuAenGluIleTyrSerTrpIleGlu 134
Db 1 GCNAGTGCACAGCTACACCAAGTACCAACAGTGGGAACGATTGAGCGGTGATTCAA 60
Qy 135 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
Db 61 CAAGTGGCCACTGATAATCCAGACCTTGTCTACTCAGAGCGTCAATTGGAACCAATTGAA 120
Qy 155 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 174
Db 121 GGACGTAAACATGTATGCTCTCAAGATT--GGTAAACCTAGACCGAATAAGCCTGCCATC 177
Qy 175 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 194
Db 178 TTCATCGATTGTGGTTTCCATTCACAGAGAGTGGATTCTCTCGCATTCGTGAGTGGTTT 237
Qy 195 Ile----- 195
Db 238 GTGAGAGAGGCTGTCGTTACCTATAATCAAGAGATCCACATGAACAGCTTCTAGATGAA 297
Qy 196 -----GlyHis----- 197
Db 298 CTGGATTCTATGTTCTCGCTGTGGTCAACATTGATGCTATGTCTACACCTGGACTAAG 357
Qy 198 AsnArgMetTrpArgLysAsnArgSerPheTyrTrpAlaAsnHisCysIleGlyThrAsp 217
Db 358 GACAGATGTGGAGAAAACCGCTCTACTATGCTGGAAGTCTCTGCTGGGTGTAGAC 417
Qy 218 LeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 237
Db 418 CCCAACAGGAATTTT---AATGCTGGCTGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474
Qy 238 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 257
Db 475 TCTGAACCTTACTGTGGACCGCCAGCTCTGAAAAGAGACAAAGGCCCTCGCAGAT 534
Qy 258 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 277
Db 535 TTCATCCCAACACCTCTCCACATCAAGCCCTACCTGACCATCCATCTATCTACAG 594
Qy 278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluLeuSer 297
Db 595 ATGATGCTCTACCTTACTCTCTATGACTACAACTGCTGAGAACTATGAGGAATTGAAT 654
Qy 298 LeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThr 317
Db 655 GCCTGTGTGAAGGTGGCGGAAGAGGCTT---GCCACTCTGCATGGCCACCAAGTACACA 711
Qy 318 HisGlyHisGlySerGluThrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 337
Db 712 TATGCCCAAGAGCTACCAACATCTATCTGCTGCTGGGGATCTGACGACTGCTCTAT 771
Qy 338 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 357
Db 772 GATCAGGGAATCAAATATTCTTTTACCCTTTTGAACCT-CCGGGATACAGCTTCTTTGGCTT 830
Qy 358 Ser 358
Db 831 TCT 833

RESULT 15

US-08-696-139-3

; Sequence 3, Application US/08696139

; Patent No. 5672496

; GENERAL INFORMATION:

APPLICANT: Fayerman, Jeffrey T.
APPLICANT: Greenen, David P.
APPLICANT: Herzhberger, Charles L.
APPLICANT: Larson, Jeffrey L.
APPLICANT: Sterner, Jane L.
APPLICANT: Zhang, Haichao
TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
PANCREATIC CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/696,139
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,258
FILING DATE: 16-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..921
US-08-696-139-3

Alignment Scores:
Pred. No.: 2,9e-62 Length: 921
Score: 576.00 Matches: 116
Percent Similarity: 59.09% Conservative: 40
Best Local Similarity: 43.94% Mismatches: 68
Query Match: 30.14% Indels: 40
DB: 1 Gaps: 5

US-09-980-881A-2 (1-360) x US-08-696-139-3 (1-921)

Qy 120 TyrGluGlnTyrHisSerLeuAenGluIleTyrSerTrpIleGluPheIleThrGluArg 139
Db 16 TATGAGAAGTACAACTACCTGGGAAACGATCGAGCTTGGACTAAGCAAGTCACCAAGTGAA 75
Qy 140 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 159
Db 76 AATCCAGACCTCATCTCTCGCAGCCATCGGAACCTACATTTTATAGAAACAATATATAC 135
Qy 160 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 179
Db 136 CTCCTCAAGGTT--GGCAAACTGGACCAATAAGCCTGCCATTTTTCATGGACTGTGGT 192
Qy 180 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 195
Db 193 TTCATGCCAGAGAATGGATTTCCTCATGCAATTTTGGCAGTGGTTTGTGAGAGAGGCTGTT 252

```
Qy 195 ----- 195
Db 253 CTCACCTATGATATGAGAGTCACATGACAGAAATTCCTCAACAAGCTAGACTTTTATGTC 312
Qy 196 -----GlyHis-----AsnArgMetTyrArg 202
Db 313 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGACCGAATGTGGAGA 372
Qy 203 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPhe 222
Db 373 AAGACCCGCTCTACCAATGCTGGAACCTACCTGCAATGGCACAGACCCCAACAGAAATTTT 432
Qy 223 ValSerLysHisTyrCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 242
Db 433 ---GATGCTGGTGGTGACAACTGGAGCCCTCTACAGACCCCTCGATGAGACTTACTGT 489
Qy 243 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 262
Db 490 GGATCTGCTGAGAGTCTGAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAC 549
Qy 263 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 282
Db 550 CTCTCCTCCATCAAAGCATACCTGACGATCCACTCATATCTCACAGATGATACTCTACCT 609
Qy 283 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 302
Db 610 TATTCCTATGATTACAAACTCCCGAGAACAAATGCTGAGTTGAATAACCTGGCTAAGGCT 669
Qy 303 AlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 322
Db 670 GCCGTGAAGAAGACTT---GCTACTGTATGGCAAGTACATACGCGCCAGGAGCT 726
Qy 323 GluThrLeuTyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLys 342
Db 727 ACAACAATCTATCTGCTGCTGGGGCTCTGATGACTGGGCTTATGACCAAGGATCAAA 785
Qy 343 TyrSerPheThr 346
Db 787 TATTCCTTCACC 798
```

Search completed: January 2, 2004, 09:25:32
Job time : 193.344 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 20:10:07 ; Search time 875.574 Seconds
(without alignments)
1423.134 Million cell updates/sec

Title: US-09-980-881A-2

Perfect score: 1911

Sequence: 1 MKLCSLAVIPVILVFCRHV.....IKYFTSNPPVKELPLSLK 360

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135919_26949/app_query.fasta_1.1628
-DB=Published Applications NA -OFMT=faeTap -SUFFIX=rnppb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -MMAP=US09980881@cgn_1_1_704 @runat_31122003_135919_26949
-NCPU=6 -ICPU=3 -NO WARN -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100
-NLOGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
------------	-------------	--------	----	-------------

1	1832	95.9	1728	10	US-09-880-107-2396	Sequence 2396, Ap
2	1729	90.5	1272	13	US-10-379-836-1	Sequence 1, Appli
3	889	46.5	1400	9	US-09-925-302-24	Sequence 24, Appl
4	623	32.6	12154	10	US-09-910-059-124	Sequence 124, App
5	621.5	32.5	1332	10	US-09-954-456-1141	Sequence 1141, Ap
6	620.5	32.5	1633	13	US-10-341-434-187	Sequence 187, App
7	619.5	32.4	1254	15	US-10-229-546-3	Sequence 3, Appli
8	619.5	32.4	1622	12	US-10-429-802-21	Sequence 21, Appli
9	619.5	32.4	1622	15	US-10-229-546-1	Sequence 1, Appli
10	611.5	32.0	1311	14	US-10-200-344-9	Sequence 9, Appli
11	611.5	32.0	1993	12	US-10-274-639-33	Sequence 33, Appl
12	591	30.9	416	10	US-09-960-352-14595	Sequence 14595, A
13	564	29.5	2128	14	US-10-200-344-13	Sequence 13, Appl
14	563	29.5	1125	9	US-09-888-615-2	Sequence 2, Appli
15	563	29.5	1332	16	US-10-176-306-75	Sequence 75, Appl
16	563	29.5	1603	16	US-10-176-306-73	Sequence 73, Appl
17	556	29.1	1826	13	US-10-252-157-453	Sequence 453, App
18	546	28.6	1200	15	US-10-300-910-7	Sequence 7, Appli
19	543	28.4	1870	10	US-09-910-059-112	Sequence 112, App
20	528.5	27.7	1050	14	US-10-200-344-11	Sequence 11, Appl
21	519.5	27.2	1311	15	US-10-200-910-5	Sequence 5, Appli
22	511.5	26.8	1311	14	US-10-076-535-3	Sequence 3, Appli
23	511.5	26.8	1687	14	US-10-076-535-1	Sequence 1, Appli
24	508	26.6	1306	9	US-09-923-779-144	Sequence 144, App
25	508	26.6	1306	9	US-09-962-436-557	Sequence 557, App
26	504	26.4	1341	9	US-09-925-297-69	Sequence 69, Appl
27	489.5	25.6	1254	9	US-09-925-297-58	Sequence 58, Appl
28	489.5	25.6	1260	13	US-10-345-680-66	Sequence 66, Appl
29	489.5	25.6	1380	10	US-09-969-347-209	Sequence 209, App
30	489.5	25.6	1380	13	US-10-345-680-64	Sequence 64, Appl
31	487	25.5	397	10	US-09-960-352-8425	Sequence 8425, Ap
32	483	25.3	2786	11	US-09-946-374-233	Sequence 233, App
33	483	25.3	2786	13	US-10-015-387A-233	Sequence 233, App
34	483	25.3	2786	13	US-10-006-130A-233	Sequence 233, App
35	483	25.3	2786	13	US-10-199-672-307	Sequence 307, App
36	483	25.3	2786	13	US-10-006-172A-233	Sequence 233, App
37	483	25.3	2786	13	US-10-187-749-307	Sequence 307, App
38	483	25.3	2786	13	US-10-194-457-307	Sequence 307, App
39	483	25.3	2786	13	US-10-184-642-307	Sequence 307, App
40	483	25.3	2786	13	US-10-196-747-307	Sequence 307, App
41	483	25.3	2786	13	US-10-015-392A-233	Sequence 233, App
42	483	25.3	2786	13	US-10-017-253A-233	Sequence 233, App
43	483	25.3	2786	13	US-10-173-689-307	Sequence 307, App
44	483	25.3	2786	13	US-10-173-690-307	Sequence 307, App
45	483	25.3	2786	13	US-10-173-691-307	Sequence 307, App

ALIGNMENTS

RESULT 1
US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Alignment Scores:
Pred. No.: 3,67e-222 Length: 1728
Score: 1832.00 Matches: 357
Percent Similarity: 86.27% Conservative: 1
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 95.87% Indels: 55
DB: 10 Gaps: 2

US-09-980-881A-2 (1-360) x US-09-880-107-2396 (1-1728)

```
Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 20 ATGAAGCTTTGACGCTTGCAGTCTCTGTACCCATTTGTTCTCTCTGTGACGACATGTC 79
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 80 TTCGGCTTTCAGAGTGGCCAAAGTTCTAGCTCTCTCTCTAGAACCTCTAGGCAAGTTCAA 139
Qy 41 ValLeuGlnAnLeuThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 60
Db 140 GTTCTACAGAATCTTACTACACATATGAGATTGTTCTCTGCGCAGCCGGTAACAGCTGAC 199
Qy 61 LeuIleValLysLysGlnValHisPhePheValAlaSerAspValAsnVal 80
Db 200 CTTATTGTGAAGAAAACAAAGTCCATTTTGTAAATGATCTGTGATGTCGACAAATGTG 259
Qy 81 LysAlaHisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAsp 100
Db 260 AAAGCCCATTTAAATGTGAGCGGAATTCATGTCAGTGTCTTGTGCGACAGCTGGAAGAT 319
Qy 101 LeuIleGlnGlnGlnIleSerAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 320 CTTATTCAACAGCAGATTTCACAGACACAGTCCAGCCCGAGCTCCGCACTCGTACTAT 379
Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerThrPheLeuPheLeuThrGluArgHis 140
Db 380 GAACAGTATCACTCACTAATGAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439
Qy 141 ProAspMetLeuThrLysIleHisLeuGlySerPheGluLysTyrProLeuTyrVal 160
Db 440 CCTGATATGCTTACAAAAATCCACATTTGATCTCTCAATTTGAGAAGTACCCACTCTATGTT 499
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAlaIleThrPheAspCysGlyIle 180
Db 500 TTAAGGTTTCTGGAAGAAGACAAACAGCCAAAAATGCCATATGATGATGCTGTGAATC 559
Qy 181 HisAlaArgGluThrPheSerProAlaPheCysLeuThrPheIleGlyHis----- 197
Db 560 CATGCCAGAAATGGATCTCTCTGCTTTCTGCTTTGCTTTGCTTTGCTTTGCTTTGCTTTG 619
Qy 197 ----- 197
Db 620 TTCTATGGATAATAGGCAATATACCAATCTCTGAGGCTTTGCTTTGCTTTGCTTTGCTTTG 679
Qy 198 -----AsnArgMetTyrArgLys 203
Db 680 CCGGTGGTTAATGTGACGGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGAGAAAG 739
Qy 204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
Db 740 AACCGTCTTCTTATGCGAAACAATCATTTGCATCGGAACAGACACCTGGAATAGGAATTTGCT 799
Qy 224 SerLysHisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 243
Db 800 TCCAAAACACTGGTGTGAGAGAGTGCATCCAGTTCTCTCATGCTCGGAACCTACTGTGGA 859
Qy 244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
Db 860 CTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGTTCTTCTTGAGAAAGAAATATC 919
```

```
Qy 264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
Db 920 AACGAGATTAAAGCATATCATCAGCATGCATTCATCATCATCCAGCATATAGTGTTCATAT 979
Qy 284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
Db 980 TCCTATATACAGAAAGTAAAGCAAGACCATGAGGAACCTGTCTCTAGTAGCCAGTGAAGCA 1039
Qy 304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
Db 1040 GTTCGTGCTATTGAGAAACCTAGTAAAAATACCAAGGTATACACATGGCCATGGCTCAGAA 1099
Qy 324 ThrLeuTyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyr 343
Db 1100 ACCTTATACCTAGTCTCTGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATAT 1159
Qy 344 SerPhe-----Th 346
Db 1160 TCGTTTAAATTGAACCTTCGAGATACGGGCACATACGATTTCTTGTCTGCCGAGCGTTAC 1219
Qy 346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
Db 1220 ATCAAAACCCACCTGTAGAGAGCTTTTTCGCGCTGTCTCTAAAA 1262
RESULT 2
US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Papio hamadryas
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1269)
US-10-379-836-1
Alignment Scores:
Pred. No.: 2,8e-209 Length: 1272
Score: 1729.00 Matches: 338
Percent Similarity: 83.37% Conservative: 8
Best Local Similarity: 81.45% Mismatches: 14
Query Match: 90.48% Indels: 55
DB: 13 Gaps: 2
US-09-980-881A-2 (1-360) x US-10-379-836-1 (1-1272)
Qy 1 MethLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 1 ATGAAGCTTTGACGCTTTCAGTCTCTGAGTCTCTTGTACCCATTTGTTCTCTCTGTGACGACATGTC 60
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
Db 61 TTCGGTTTTCAGAGTGGCCAGGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAA 120
Qy 41 ValLeuGlnAnLeuThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 60
Db 121 GTGTACAGAAATCTTACTACAAATATGAGATTGTTCTTGGCAGCCGGTACAGCGGAC 180
Qy 61 LeuIleValLysLysGlnValHisPhePheValAlaSerAspValAsnVal 80
Db 181 CTTATTGAGAAAGAAACAAAGTCCATTTTGTAAATTCATCTGTGACAAATGTG 240
```


Db 610 CAATATTGCTTTTACAAATTGAACCTTCGAGATACGGGCACATACGAGTTCCTTGTGCGCGA 669
Qy 346 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLys 360
Db 670 CGGTTCATCAACCCACCTCTAGAGAGCTTTTGGCGCTGCTCTCTAAAA 719

RESULT 4

US-09-910-059-124
; Sequence 124, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; FILE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; PRIOR FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 124
; LENGTH: 2154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-124

Alignment Scores:
Pred. No.: 3,77e-68 Length: 2154
Score: 623.00 Matches: 132
Percent Similarity: 53.40% Conservative: 72
Best Local Similarity: 34.55% Mismatches: 134
Query Match: 32.60% Indels: 44
DB: 10 Gaps: 7

US-09-980-881A-2 (1-360) x US-09-910-059-124 (1-2154)

Qy 3 LeuCySerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
Db 1 ATGTGGCAGCTCTGTTCTGGTACTGTGGCCCTGGCATCTGCTCATCATGTTGGTGAG 60
Qy 23 ---PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41
Db 61 CACTTTGAAGCGGAGAGGTGTCGTGTTAAGCTTGAAGATGAATAATCAATTAACATA 120
Qy 42 LeuGlnAsnLeuThrThrThrGluLeuValLeuTyrGlnProValThrAlaAspLeu 61
Db 121 ATCCGCGAGTTGGCCAGCAGCAGCCAGATTCCTGGAAGCCAGATTCCTGCACACAA 180
Qy 62 IleValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLys 81
Db 181 ATCAAACTCAGTACAGTTCCTCGTGTAAAGCAGAGATCTGCTCAGTGTGAG 240
Qy 82 AlaHisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeu 101
Db 241 AATGTTCTAAAGCAGAGTGAACATAAATCAAGGTACTGATAAGCAACCTGAGAAATGTG 300
Qy 102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrGlu 121
Db 301 GTGGAGGCTCAGTTGATAGCGGGT-----CGTGAACAGGACAGATTATGAG 351
Qy 122 GlnTyrHisSerLeuAsnGluIleThrSerTrpIleGluPheIleThrGluArgHisPro 141

Db 352 AAGTACAACAAGTGGGAAACGATAGAGCTTCGACTCAACAAGTCGCCACTGAGATCCA 411
Qy 142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
Db 412 GCCCTCATCTCTCGCAGTGTATCGGAACACCATTTGAGGGACCGCTATTATACCTCTG 471
Qy 162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
Db 472 AAGGTT---GGCAAGCTGGACAAAATAAGCTCCCATTTTCATGGAGTGGTTTCAT 528
Qy 182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 195
Db 529 GCCAGAGAGTGGATTCTCTCGCATTCGCCAGTGGTTTGTAAAGAGAGGCTGTTGCTACC 588
Qy 195 ----- 195
Db 589 TATGGAGCTGAGATCCAAAGTCAGACAGCTTCTCGCAAGTTAGACTTTTATGTCTGCT 648
Qy 196 -----GlyHis-----AsnArgMetTTPATGlyAsn 204
Db 649 GTGCTCAATATTGATGGCTACATCTACCTGGACCAAGAGCCGATTTTGGAGAAAGACT 708
Qy 205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
Db 709 CGCTCCACCCATCTGATCTAGCTGCTGGCACAGACCCCAACAGAAATTT---GAT 765
Qy 225 LysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeu 244
Db 766 GCTGGTGGTGTGAATTTGGAGCCTCTCGAAACCCCTGTGATGAATTAATCTACTGTG 825
Qy 245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIleAsn 264
Db 826 GCCGACAGCTGAAAGAGACCAAGGCCCTGGCTGATTTTCATCCGCAACAAACTCTCT 885
Qy 265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 284
Db 886 TCCATCAAGGCATATCTGACATCCACTCTGCTACTCCCAATGATGATCTACCTTACTCA 945
Qy 285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
Db 946 TATGCTTACAAACTCGCGTGAGAACATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTG 1005
Qy 305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
Db 1006 AAGAAGCTT---GCCTCATCTGACGCGCACCAAGTACACATATGCGCGGAGCTACACA 1062
Qy 325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
Db 1063 ATCTATCTCTCTGCTGGGACTTCTAAAGACTGGGCTTATGACCAAGGAATCAGATATCC 1122
Qy 345 PheThr 346
Db 1123 TTCACC 1128

RESULT 5

US-09-954-456-1141
; Sequence 1141, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134

[illegible]

```

Qy 324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleIysTyr 343
Db 1071 ACAATTACCGGATATACAGGTTCTCTTTAGACTGGGCTTATGACCTGGGCATCAACAC 1130
Qy 344 SerPhe 345
Db 1131 ACAATT 1136

RESULT 7
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299. A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MP010-156PIRNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-229-546-3

Alignment Scores:
Pred. No.: 4,22e-68 Length: 1254
Score: 619.50 Matches: 130
Percent Similarity: 54.07% Conservative: 76
Best Local Similarity: 34.12% Mismatches: 128
Query Match: 32.42% Indels: 47
DB: 15 Gaps: 8

US-09-990-881A-2 (1-360) x US-10-229-546-3 (1-1254)
Qy 6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla----- 22
Db 1 ATGAGGCTCATCTGCTGCTGGGTTTGATGTGTACCACTTGCATTCCTGCTCTGTCGCG 60
Qy 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
Db 61 TTTGACAGGAGAAAGGTGTTCCGCGTGAAGCCCGAGGATGAAAAACAAGCAGACATCAT 120
Qy 43 GlnAsnLeuThrThrTyrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 62
Db 121 AAGGACTTGCCCAAAACCAATGAGTTGACTTCTGGTATCCAGGTCACCCACCAACGTA 180
Qy 63 ValIysIysIysGlnValHisPhePheValAsnAlaSerAspValAspAsnValIysAla 82
Db 181 GCTGCTAAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCT 240
Qy 83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 102
Db 241 GCCTTGATCAAAATAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATT 300
Qy 103 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
Db 301 GAGAAACAGATTGTGTTAAAGAAGATATCCCAAGGAGGCACAGC-----TAGCA 351
Qy 122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
Db 352 AAATACAAATAATTTGGGAAAAGATTTGGCTTTGGACTGAAGAAGATGATGGTAAGTATCT 411
Qy 142 AspMetLeuThrIysIleHisIleGlySerSerPheGluIysTyrProLeuTyrValLeu 161

```

```
Db 412 GAAATGGCTCTCGTATTAAATTTGATCTACTGTTGAAGATAATCCACTATATGTTCTG 471
Qy 162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
Db 472 AAGATT---GGGAAAGAAATCAAGAAAGAAAGGCTATTTTATGATGTTGGCATTCAC 528
Qy 182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPhele----- 195
Db 529 GCACGAGAATGGTCTCCCGACATTTCTGCAGTGGTTTGTCTATCAGGCAACCAAAACT 588
Qy 195 ----- 195
Db 589 TATGGGAGAAACAAATATATGACCAAACTTTGGACCGAATGAATTTTATCTTCTCT 648
Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsn 204
Db 649 GTGTTCAATGTTGATGATATATTGTTGTCATGGACAAAGACCGCATGTGGAGAAAT 708
Qy 205 ArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
Db 709 CGTTCCAAAGAACCAAACTCCAAATGCAATCGGCACTGACCTCAACAGGAATTTT---AAT 765
Qy 225 LysHisTrpCysGluGluGlyAlaSerSerSerSerSerSerGluThrTyrCysGlyLeu 244
Db 766 GCTTCATGGAATCCATTCCTTAACCAACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825
Qy 245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
Db 826 GCACCAAGTTCGAGAAAGACGAAAGCTGTCTACTAATTTTATAGAACCACTGAT 885
Qy 265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 284
Db 886 GAAATCAAGGTTTACATCACCCTTCATCTCTACTCCAGATGCTATTGTTTCCCTATGGA 945
Qy 285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
Db 946 TATACATCAAAATGCCACCTTAACCATCAGAGACTTGGCCAAAGTTGCCAAGATTGGCACT 1005
Qy 305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
Db 1006 GATGTTCTA---TCAACTCGATATGAACCCGCTACATCTATGCGCCCAATAGATCAACA 1062
Qy 325 LeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
Db 1063 ATTTACCGCATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAACACACA 1122
Qy 345 Phe 345
Db 1123 TTT 1125
```

RESULT 8

```
US-10-429-802-21
; Sequence 21, Application US/10429802
; Publication No. US2003028285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: WONG, KA YIN
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-21
```

```
Alignment Scores:
Pred. No.: 6.51e-68 Length: 1622
Score: 619.50 Matches: 130
Percent Similarity: 54.07% Conservative: 76
Best Local Similarity: 34.12% Mismatches: 128
Query Match: 32.42% Indels: 47
DB: 12 Gaps: 8

US-09-980-881A-2 (1-360) x US-10-429-802-21 (1-1622)

Qy 6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla----- 22
Db 1 ATGAGGCTCATCTCGCTGCTGGTGGTTGATTCCTACCCTCTTGCATTTGCTCTGCTCGC 60
Qy 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
Db 61 TTTTCACAGGAGAGAGGTTTCCCGGTGAAGCCCGAGGATGAAGAAACAGCAGACATCATA 120
Qy 43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 62
Db 121 AAGGACTTGGCCAAACCAATGAGCTTGAATCTTCTGTATCCAGGTGCCACCCACCGTA 180
Qy 63 ValLysLysGlnValHisPheValAsnAlaSerAspValAspValValValAla 82
Db 181 GCTCCTAATATGATGGTGGATTTCCGAGTTAGTGAGAGGAATCCCAAGCCATCCAGTCT 240
Qy 83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 102
Db 241 GCCTTGGATCAATAAATGCATCTATCAATCTTGTATTCATGATCTACAAGAGAGATT 300
Qy 103 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
Db 301 GAGAAACAGTTTGTATGTTAAAGAGATATCCAGCGAGGACAGC-----TACGCA 351
Qy 122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
Db 352 AATACAAATATTCGGAAAAGATTGTGGCTGGACTGAAAAGATGATGGATTAAGTATCCT 411
Qy 142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
Db 412 GAATGCTCTCTGTTATTAATTTGGATCTACTGTTGAAGATATATCACTATATGTTCTG 471
Qy 162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
Db 472 AAGATT---GGGAAAGAAATGAAGAAAGAGCTATTTTATGATGTTGGCATTCAC 528
Qy 182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPhele----- 195
Db 529 GCACGAGAATGGTCTCCCGACATTTCTGCAGTGGTTTGTCTATCAGGCAACCAAAACT 588
Qy 195 ----- 195
Db 589 TATGGGAGAAACAAATATATGACCAAACTTTGGACCGAATGAATTTTATCTTCTCT 648
Qy 196 -----GlyHis-----AsnArgMetTrpArgLysAsn 204
Db 649 GTGTTCAATGTTGATGATATATTGTTGTCATGGACAAAGACCGCATGTGGAGAAAT 708
Qy 205 ArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
Db 709 CGTTCCAAAGAACCAAACTCCAAATGCAATCGGCACTGACCTCAACAGGAATTTT---AAT 765
Qy 225 LysHisTrpCysGluGluGlyAlaSerSerSerSerSerSerGluThrTyrCysGlyLeu 244
Db 766 GCTTCATGGAATCCATTCCTTAACCAACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825
Qy 245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
Db 826 GCACCAAGTTCGAGAAAGACGAAAGCTGTCTACTAATTTTATAGAACCACTGAT 885
Qy 265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 284
Db 886 GAAATCAAGGTTTACATCACCCTTCATCTCTACTCCAGATGCTATTGTTTCCCTATGGA 945
```


742	ACCAATGATCGATTTTGGAGAAAAACAAGGTCAAGGAACCTCAAGGTTTTCCTGCCGTGGA	801
216	ThrAspLeuAsnSerAsnPheValSerLysHisTyrCysGluGluGlyAlaSerSerSer	235
802	GTGGATGCCAATAGAACTGGAAAGTGAG---TGGTGTGATGAGAGAGCTTCTATGCAC	858
236	SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVallyAlaVal	255
859	CCTTGTGTGATGACACATCTGTGGCCCTTTTCCAGAAATCTGAGCCGGAAGTGAAGGCTGTA	918
256	AlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyr	275
919	GCTAACTTCTCTCGAAAAACAGAAACGACATTAGGGCTTATCTCTCTTCATGCATAT	978
276	SerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGlu	295
979	GCTCAGATGTTACTGTATCCCTATCTTCAATAATGCAACAATCCCAATTTTATAGTGT	1038
296	LeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArg	315
1039	GTGGAATCTCGACCTTATAAAGCTGTGAATGCACATT---CAGTCAGTATACGGGTACGA	1095
316	TyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTyr	335
1096	TACAGATATGGACCGCCCTCCACACGTTGTATGTGAGCTCTGGTAGCTCAATGGGATTGG	1155
336	IleTyrAspLeuGlyIleLysTyrSerPhe-----	345
1156	GCTACAAAATGGAAATACCTTATGCAATTTGCTTTCGAACCTAGTCAGACACTGGATATTTT	1215
346	-----ThrSerAsnProProValGluLysLeu	354
1216	GGATTTTACTCCAGAGATGCTCATCAAAACCCACCTGTACAGAAACTA	1264

RESULT 11

US-10-274-639-33

; Sequence 33, Application US/10274639

; Publication No. US20030232349A1

; GENERAL INFORMATION:

; APPLICANT: INCYTE GENOMICS, INC.

; APPLICANT: DELEGEANE, Angelo M.; GANDHI, Ameena R.

; APPLICANT: HAFALIA, April J.A.; LU, Dyung Aina M.

; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.

; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.

; APPLICANT: NGUYEN, Dannie B.; LEE, Ernestine A.

; APPLICANT: KHAN, Farrah A.; YUE, Henry

; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.

; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi

; APPLICANT: YANG, Junming; THANGAVELU, Kavitha

; APPLICANT: DING, Li; KEARNEY, Liam

; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.

; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.

; APPLICANT: BURFORD, Neil; WALIA, Narinder K.

; APPLICANT: LAL, Preeti G.; LEE, Sally

; APPLICANT: TODD, Stephen; LO, Terence P.

; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.

; APPLICANT: AZIMZAI, Valda; LU, Yan

; TITLE OF INVENTION: PROTEASES

; FILE REFERENCE: PI-0167 USA

; CURRENT APPLICATION NUMBER: US/10/274,639

; CURRENT FILING DATE: 2002-10-18

; PRIOR APPLICATION NUMBER: PCT/US01/22397

; PRIOR FILING DATE: 2001-07-17

; PRIOR APPLICATION NUMBER: US 60/220,063

; PRIOR FILING DATE: 2000-07-21

; PRIOR APPLICATION NUMBER: US 60/221,680

; PRIOR FILING DATE: 2000-07-28

; PRIOR APPLICATION NUMBER: US 60/223,544

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: US 60/224,717

; PRIOR FILING DATE: 2000-08-11

; PRIOR APPLICATION NUMBER: US 60/225,988

; PRIOR FILING DATE: 2000-08-16

;; PRIOR APPLICATION NUMBER: US 60/227,568

;; FILING DATE: 2000-08-23

;; NUMBER OF SEQ ID NOS: 42

;; SOFTWARE: PERL Program

;; SEQ ID NO 33

;; LENGTH: 1993

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; NAME/KEY: misc.feature

;; OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CB1

US-10-274-639-33

Alignment Scores:

Pred. No.: 9,628-67 Length: 1993
Score: 611.50 Matches: 141
Percent Similarity: 50.36% Conservativity: 69
Best Local Similarity: 33.81% Mismatches: 135
Query Match: 32.00% Indels: 72
DB: 12 Gaps: 11

US-09-980-881A-2 (1-360) x US-10-274-639-33 (1-1993)

Qy 1 MetLysLeuCys-----SerLeuAlaValLeuValProLleValLeuPheCysGluGln 18
Db 380 CTGCTCTTTGCTGCTCTTTTGAAGATTCTGCAACCG-----GGGCACAGC 427
Qy 19 HisValPhe-----AlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSer 36
Db 428 CACCTTTTAAACAACCGCTGCTGCTGATAAAGTGATAAGATTATTATCCCAAAACAGAA 487
Qy 37 ArgGlnValGlnValLeuGlnAsnLeuThrThrThrThrThrThrThrThrThrThrThr 56
Db 488 GAGGAAGCATATGCACTGAGAAATATCTTATCAACTTAAGGTGAGCTGTGGAGCGCC 547
Qy 57 ValThrAlaAspLeuLeuValLys-----LysLysGlnValHisPheValAsnAla 74
Db 548 AGCAGTATCTCTGATGATCAGAGGACAGTACTGATGTCATATCCCAAAATGGT 607
Qy 75 SerAspValAspAsnValLysAlaHisLeuAsnValSerGlyLeuProCysSerValLeu 94
Db 608 TCCCGAGCC-----CTGTTAGCTCTTACAGGAAGCAACATCCAGTCAAGGTCTCTC 661
Qy 95 LeuAlaAspValGluAspLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 114
Db 662 ATAGAAGATCTTCAGAAACACTGGAGGAAGGAGCAGCTTGCACACCAGCAAGAACCGA 721
Qy 115 AlaSerAlaSer-----TyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrp 132
Db 722 AGATCCCTCTGATATATTAATGAGTATCTCTCTTAGAAGAAATTCAAAATGG 781
Qy 133 IleGluPheLeuThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSer 152
Db 782 ATGCATCATCTGAATAAAACTCACTCAGGCTCATTCACATGTTCTCTATTGGAAGATCA 841
Qy 153 PheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 172
Db 842 TATGAGGGAAGATCTCTTTTATTATAAGCTG---GGCAGACGATCAGCTCAAAAGA 898
Qy 173 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeu 192
Db 899 GCTGTTGGATAGACTGTGGTATTCATCGAAGAGNATGGATTGGTCTGCTCTTTTGTCAG 958
Qy 193 TrpPheIle----- 195
Db 959 TGGTTTGTAAAGAAGCTCTTCTAACATATAAGAGTGACCCAGCCATGAGAAATGTTG 1018
Qy 195 ----- 195
Db 1019 AATCATCTATTTCTATATCANGCTGTGTTTAACGTCGATGGATACCATTTTAGTTGG 1078
Qy 196 GlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGly 215
Db 196 ----- 215

Db 1079 ACCAATGATCGATTTTGGAGAAAAACAAGGTCAGGAACCTCAAGGTTTCGCTGCGGTGA 1138
Qy 216 ThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSer 235
Db 1139 GTGGATGCCAATAGAAACTGGAAAGTGAAG---TGGTGTGATGAAGAGCTTCTATGCAC 1195
Qy 236 SerCysSerGluThrTyrCysGlyLeuTyrProGlnSerGluProGluValLysAlaVal 255
Db 1196 CTTGTGATGACATATCTGTGGCCCTTTCCAGATCTGAGCGGAGTGAAGGCTGTA 1255
Qy 256 AlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyr 275
Db 1256 GCTAACTTCTTCGAAAAACACAGAAAGCACATTAGGCTTATCTCTCTTTCATGCATAT 1315
Qy 276 SerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGlu 295
Db 1316 GCTCAGATGTTACTGTATCTCTTATTCACAAATATGCAACAAATTCCTCAATGATGT 1375
Qy 296 LeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArg 315
Db 1376 GTGGAATCTGAGCTTATAAGCTGTGAATGCACTT---CAGTCGATATACGGGTACGA 1432
Qy 316 TyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrp 335
Db 1433 TACAGATATGACGAGCTCCACACGCTTGTATGTAGCTCTGCTAGCTCAATGATGG 1492
Qy 336 IleTyrAspLeuGlyIleLysTyrSerPhe----- 345
Db 1493 GCCTACAAAAATGGAATACCTTATGCAATTTGCTTTCGAACTACGTGACACTGGATATTT 1552
Qy 346 -----ThrSerAsnProValGluLysLeu 354
Db 1553 GGATTTTACTCCAGAGATGCTCATCAACCCACCTGTACAGAACTA 1601

RESULT 12

US-09-960-352-14595

;; Sequence 14595, Application US/09960352

;; Patent No. US20020137139A1

;; GENERAL INFORMATION: Wesley C.

;; APPLICANT: Warren, Wesley C.

;; APPLICANT: Tao, Nengbing

;; APPLICANT: Byatt, John C.

;; APPLICANT: Mathialagan, Nagappan

;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

;; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

;; FILE REFERENCE: 16511.006/37-21(10298)C

;; CURRENT APPLICATION NUMBER: US/09/960,352

;; CURRENT FILING DATE: 2001-09-24

;; NUMBER OF SEQ ID NOS: 15112

;; SEQ ID NO 14595

;; LENGTH: 416

;; TYPE: DNA

;; ORGANISM: Bos taurus

;; OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10

US-09-960-352-14595

Alignment Scores:

Pred. No.: 2,76e-65 Length: 416
Score: 591.00 Matches: 111
Percent Similarity: 89.13% Conservativity: 12
Best Local Similarity: 80.43% Mismatches: 15
Query Match: 30.93% Indels: 0
DB: 10 Gaps: 0

US-09-980-881A-2 (1-360) x US-09-960-352-14595 (1-416)

Qy 51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGlnValHisPhe 70
Db 2 ATTTCTCTGGCAGCAGTAGCAGCTGAATATATTGTGAAGGATAGAGTCAATTTT 61
Qy 71 PheValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIlePro 90
Db 62 TTTGTGAATGATCTGATGAAGCAATGTGAAGCCCAATTTAAATGCGAGCAAGATCCA 121

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:47:17 ; Search time 3188.89 Seconds
(without alignments)
2743.777 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLPVILFCEQHV.....IKYFTSNPPVEKLLPLSLK 360

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool/US09980881/runat 31122003 135916 26873/app query.fasta_1.1628
-DB-EST -QFMT=fastap -SUPFIX=ret -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN 1 1 5167 @runat 31122003 135916 26873 -NCPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1534.5	80.3	1446	11	AK004045	AK004045 Mus muscu
2	1351.5	70.7	971	9	AL536069	AL536069 ALS36069
3	1034	54.1	788	10	BG618629	BG618629 603646064
4	1016	53.2	702	2	HSM08979	Bx466754 Homo sapi
5	1015.5	53.1	853	12	B1144644	B1144644 602909932
6	996.5	52.1	738	10	BG616456	BG616456 602614618
7	989	51.8	1004	14	BY704293	BY704293 BY704293
8	983	51.4	747	10	BG566051	BG566051 602584652
9	975	51.0	725	10	BG618239	BG618239 603645967
10	965	50.5	852	10	BG569281	BG569281 602588573
11	957.5	50.1	1022	12	B1332659	B1332659 602984408
12	949	49.7	735	9	AV646979	AV646979 AV646979
13	927.5	48.5	750	10	BG618813	BG618813 602646186
14	921	48.2	541	14	CB157619	CB157619 K-EST0216
15	902	47.2	747	10	BG565100	BG565100 602583720
16	898	47.0	691	10	BG568240	BG568240 602587149
17	894.5	46.3	724	9	A1048167	A1048167 ud71b12.y
18	868	45.4	583	9	AV693037	AV693037 AV693037
19	857	44.8	602	9	AV692032	AV692032 AV692032
20	849	44.4	601	9	AV658390	AV658390 AV658390
21	848	44.4	623	9	AV694058	AV694058 AV694058
22	838	43.9	621	9	AV698425	AV698425 AV698425
23	837.5	43.8	749	10	BG617911	BG617911 602645566
24	826	43.2	485	13	EX091627	EX091627 BX091627
25	818.5	42.8	888	9	AV651709	AV651709 AV651709
26	815.5	42.7	806	10	BG567264	BG567264 602589745
27	814.5	42.6	577	14	CB154355	CB154355 K-EST0212
28	806.5	42.2	661	9	A1529872	A1529872 ui83d11.y
29	806	42.2	580	9	A1182113	A1182113 ud73f10.y
30	806	42.2	589	9	AA968307	AA968307 uc71h07.y
31	803.5	42.0	603	10	AW950700	AW950700 EST362770
C	799.5	41.8	973	9	AL536068	AL536068 ALS36068
32	782	40.9	510	14	CB156851	CB156851 K-EST0215
33	773.5	40.5	537	14	N98450	N98450 za28c09.r1
34	767	40.1	958	9	A1322676	A1322676 mj24d06.y
35	766	40.1	511	9	A1574117	A1574117 uj68d09.y
36	765.5	40.1	787	14	CB593837	CB593837 AGENCOURT
37	759	39.7	511	9	A1528146	A1528146 ui94a04.y
38	742	38.8	462	9	AV681868	AV681868 AV681868
C	741.5	38.8	843	10	BG216526	BG216526 RST36106
40	731	38.3	455	14	W88434	W88434 zh69f04.r1
41	728.5	38.1	623	9	AV658178	AV658178 AV658178
42	719	37.6	434	14	CB164014	CB164014 K-EST0225
43	718	37.6	615	9	AV658128	AV658128 AV658128
44	711.5	37.2	558	4	BX528015	BX528015 RZPD Mus
45						

ALIGNMENTS

RESULT 1
AK004045
LOCUS
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length
enriched library, clone:1110032P04 product:carboxypeptidase B2
(plasma), full insert sequence.
1446 bp mRNA linear HTC 05-DEC-2002
AK004045
ACCESSION AK004045.1 GI:12835067
VERSION AK004045.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kaekawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischnmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuohli, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Hill, D., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Lyons, P., Mammucioni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, C., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weltz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660
 PUBMED 11217851

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 12117851

6 (bases 1 to 1446)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurthara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCAACCTCGAGTTTCTTTTCTTWN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCTCAAGAGCTCAATTAATTAACCCCTCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES
 Location/Qualifiers
 1..1446
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:1110032P04"
 /db_xref="MGI:1900711"
 /db_xref="taxon:10090"
 /clone="1110032P04"
 /tissue_type="whole body"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="18-day embryo"
 42..1310
 /note="unnamed protein product; carboxypeptidase B2 (plasma) [MGI:1891837, GB|NM_019775, evidence: BLASTN, 99%, match=1419]
 putative"
 /codon_start=1
 /protein_id="BAB231141.1"
 /db_xref="GI:12835068"
 /db_xref="MGI:1891837"
 /translation="MKHGGILVAILIYEQHGFQFQGVLSALPRTSRVQLLQNL TTTVEVLPVPTAEFTKEKVEFFVNSADSVSKAHNLNRPFPNVLNNVEDLIE QVNDFTVSPRASAYVEQYHNEIYSWIEVIYEQHDMLOKLIYIGSSPFKPYLYVI KVSCEORIKNAIWDGCIHAREWISPAFCLEWFGYVTPQHKENLYTRLLRHVDYVI MPNNVGDYDTWKNNRWRKNSRAHKNRCVGTDLNENPASKHCKGSSCSET YCGYPSSEPEKAVADFLNRNDIHKAYISMHSYSQILPFYSNRSKSKDHELSL VASAVRAIESINKNTRYTHGSGESLYLAPGSGDDWIYDIGIKYSFTIELDRTGRYG FLPERYIKPTCAEALAAISKIVHWIRNT"

BASE COUNT 421 a 329 c 311 g 385 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9.3e-165 Length: 1446
 Score: 1534.50 Matches: 302
 Percent Similarity: 78.80% Conservative: 25
 Best Local Similarity: 72.77% Mismatches: 32
 Query Match: 80.30% Indels: 56
 Gaps: 3

US-09-980-881A-2 (1-360) x AK004045 (1-1446)

Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
 Db 42 ATGAGCTTCATGGCTTGGTAATCTGTGATGCCATCATCTCTATF--GAGAGCATGGC 98
 Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
 Db 99 TTCGCCITTCAGATGGCCAGGTTTATCTCTCTCTCCAGAACCTCCAGCAAGTTCAA 158
 Qy 41 ValLeuGlnAsnLeuThrThrThrTyrGluLeuValLeuTrpGlnProValThrAlaAsp 60
 Db 159 CTACTTCAGATCTTACTACACGATGAGTGGTCTCTCTCGCAGCCAGTCAGCTGAA 218


```
Db 308 GCCCATTTAAATGTGAGCGGAATTCATGTCAGTGTCTTCTGGCAGACGTGGAAGATCTT 367
Qy 102 ILeGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrGlu 121
Db 368 ATTCAACAGCAGATTTCCAAAGACACAGTCAGCCCGAGCTCCGATCGTACTATGAA 427
Qy 122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
Db 428 CAGTATCACTACTAAATGAATCTATTCTTGGATAGATTTATAACTGAGAGCACTCT 487
Qy 142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
Db 488 GATATGCTTACAAAAATCCATTTGGATGCTCTCATTTGAGAAAGTAGTCCCACTCTATGTTTA 547
Qy 162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
Db 548 AAGGTTTCTGGAAAGAACCAAGCAGCCAAATAATGCCATATGATGATGCTGGAAATCCAT 607
Qy 182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 608 GCCAGAGAAATGGAATCTCTCTGCTTTCGCTTGTGTTGATAGGCCATATACTCAATTC 667
Qy 197 ----- 197
Db 668 TATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATGCGG 727
Qy 198 -----AsnArgMetTrpArgLysAsn 204
Db 728 GTGGTTAATGTGGATGTTATGACTACTCATGGAAGAAATCGAATGTGGAGAAAGAAC 787
Qy 205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
Db 788 CGTCTTTCTATGCGAACAAATCATTTGCAATGGAAACAGACCTGAATAGGAATTTGCTTCC 847
Qy 225 LysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeu 244
Db 848 AAACACTGGTGTGAGGAAGGTGCATCCAGTCTCTCATGCTCGGAACCTACTGTGGACTT 907
Qy 245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
Db 908 TATCTGTAGTACAGAACCAAGATGAAGCAGTGGCTAGTCTTTCTTGAGAGAAATMAAC 967

RESULT 3
LOCUS BG618629 788 bp mRNA linear EST 18-APR-2001
DEFINITION 602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
mRNA sequence.
ACCESSION BG618629
VERSION BG618629.1 GI:13670000
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 788)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM61630 row: 0 column: 15
High quality sequence stop: 760.
Location/Qualifiers
1..788
/organism="Homo sapiens"
/mol_type="mRNA"
```

FEATURES

source

```
/db xref="taxon:9606"
/clone="IMAGE:4767974"
/lab host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattatggcc); Site 2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGGCGGCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
```

BASE COUNT 222 a 157 c 169 g 240 t
ORIGIN

Alignment Scores:
Pred. No.: 9,42e-108 Length: 788
Score: 1034.00 Matches: 201
Percent Similarity: 83.82% Conservative: 1
Best Local Similarity: 83.40% Mismatches: 1
Query Match: 54.11% Indels: 38
DB: 10 Gaps: 1

US-09-980-881A-2 (1-360) x BG618629 (1-788)

Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
Db 66 ATGAAGCTTTGTCAGCTTGCAGCTCTTGTATCCCAATTTCTCTTCTGTGACGACATGTC 125
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
Db 126 TTCGCGTTTCAGAGTGGCCAGTTCTAGCTCTCTTCTAGAACCTCTAGCAAGTTTCA 185
Qy 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
Db 186 GTTCTACAGATCTTACTACACATATGAGATTTCTCTGGCAGCCGGTAACAGCTGAC 245
Qy 61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAsnVal 80
Db 246 CTTATTGTGAAGAAAAACAAAGTCCATTTTTTTGTAATGTCATCTGATGTCGACATGTG 305
Qy 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
Db 306 AAAGCCCATTTAAATGTAGCGGAATTCATGCGAGTGTCTTCTGGCAGATGTGGAAGAT 365
Qy 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db 366 CTTATTCAACAGCAGATTTCCACAGACACAGTCAGCCCGAGCTCCGCACTCGTACTAT 425
Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
Db 426 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAAATCTGAGAGGAT 485
Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
Db 486 CCTGATATGCTTACAAAAATCCATTTGATGATTCCTCATTTGAGAAAGTACCCACTCTATGT 545
Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
Db 546 TTAAGGTTTCTGGAAAGAACCAAGCAGCCAAATAATGCCATATGATGATGCTGGAAATC 605
Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
Db 606 CATGCCAGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGGCCATATAACTCAA 665
Qy 197 ----- 197
Db 666 TTCTATGGGATAATAGGGCAATATATACCAATCTCTCTGAGGCTTGTGGATTTCTATGTTAT 725
Qy 198 -----AsnArgMetTrpArg
|||||:|||||

```
Db 726 GCCAGTGGTTATGTTGGATGGTTGACTACTCATGGAAGAAATCGATTGTGGAGA 785
Qy 203 Lys 203
Db 786 AAG 788

RESULT 4
HSM089979
ID HSM089979 standard; RNA; EST; 702 BP.
XX
AC BX496754;
SV BX496754.1
XX
DT 09-MAY-2003 (Rel. 75, Created)
DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
XX Homo sapiens mRNA; EST DKFZp77911529_r1 (from clone DKFZp77911529)
XX
KW EST; expressed sequence tag.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
XX [1]
RN 1-702
RP 1-702
RA Ansoeger W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RT ; Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
XX This is the 5' sequence of the clone insert
CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC Sequenced by EMBL (European Molecular Biology Laboratories,
CC Heidelberg/Germany) within the cDNA sequencing consortium of
CC the German Genome Project.
CC No sl sequence available.
CC This clone (DKFZp77911529) is available at the RZPD in Berlin.
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
XX Key Location/Qualifiers
FH 1. 702
FT source
FT /db_xref="taxon:9606"
FT /mol_type="mRNA"
FT /organism="Homo sapiens"
FT /clone="DKFZp77911529"
FT /clone_lib="779 (synonym: hnccl). Vector pSport1_Sfi; host
FT DH10B; sites SfiIA + SfiIB"
FT /dev_stages="fetal"
FT /tissue_type="liver"
XX
SQ Sequence 702 BP; 197 A; 149 C; 146 G; 210 T; 0 other;

Alignment Scores:
Pred. No.: 9.12e-106 Length: 702
Score: 1016.00 Matches: 196
Percent Similarity: 98.99% Conservative: 0
Best Local Similarity: 98.99% Mismatches: 1
Query Match: 53.17% Indels: 1
DB: 2 Gaps: 0

US-09-980-881A-2 (1-360) x HSM089979 (1-702)

Qy 1 MetLysLeuCySerLeuAlaValLeuValProLeuValLeuPheCysGluGlnHisVal 20
Db 59 ATGAAGCTTTGGAGCGCTTGACGCTCTGTGACCATGTTCTCTCTGAGCAGCATGTC 118
Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
```

```
Db 119 TTCGGTTTCAGAGTGGCCAAAGTTCTAGCTCTCTTCTAGAACCTTAGCAAGTTCAA 178
Qy 41 ValLeuGlnAenLeuThrThrThrTyTyrGluileValLeuTrpGlnProValThrAlaAse 60
Db 179 GTTCTACAGATCTTACTACCAATATGAGATTGTTCTCTGCGAGCCGGTAACAGCTGAC 238
Qy 61 LeuileValLys-LysLysGlnValHisPhePheValAasnAlaSerAapValAapAenVa 80
Db 239 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATCGACAATGT 298
Qy 80 LysAlaHisLeuAenValSerGlyIleProCySerValLeuLeuAlaAapValGluAs 100
Db 299 GAAAGCCCATTTAAATGTGAGCGGAATTCATGAGTGTCTTGTGGCAGACGTGGAAGA 358
Qy 100 pLeuileGlnGlnIleSerAasnAapThrValSerProArgAlaSerAlaSerTyTy 120
Db 359 TCTTATTCAACAGCAGATTTCCAAACGACACAGTCAGCCCCCGAGCCTCGCATCTACTA 418
Qy 120 rGluGlnTyHisSerLeuAenGluIleTySerTrpIleGluPheIleThrGluArgHi 140
Db 419 TGAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAACTGAGAGCA 478
Qy 140 sProAseMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeuTyVa 160
Db 479 TCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCACCTCTATGT 538
Qy 160 LLeuLysValSerGlyLysGluGlnThrAlaLysAasnAlaIleTrpIleAapCysGlyI 180
Db 539 TTTAAAGTTTCTGGAAGAAAGCAAGCAGCAGCAAAAATGCCATATGGATTGACTGTGGAT 598
Qy 180 eHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
Db 599 CCATGCCAGAGAATGGATCTCTCTGCTTTCTGCTTGTGGTTTCATAGCCAT 650

RESULT 5
B1144644
LOCUS B1144644
DEFINITION B1144644 853 bp mRNA linear EST 05-JUL-2001
602909932P1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5051358 5',
mRNA sequence.
ACCESSION B1144644
VERSION B1144644.1 GI:14604645
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 853)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1139 row: 0 column: 07
High quality sequence stop: 849.
FEATURES
Location/Qualifiers
1. 853
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5051358"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
note="Organ: liver; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
```


Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT
ORIGIN

243 a 199 c 194 g 217 t

Alignment Scores:

Pred. No.: 1.41e-105 Length: 853
Score: 1015.50 Matches: 202
Percent Similarity: 76.6% Conservativity: 18
Best Local Similarity: 70.38% Mismatches: 26
Query Match: 53.14% Indels: 41
DB: 12 Gaps: 3

US-09-980-881A-2 (1-360) x B114464 (1-853)

```
QY 3 LeuCysSerLeuAlaValLeuValProLeuValLeuPheCysGluGlnHisValPheAla 22
DB 2 CTTTCATGGCTTGGAACTCTGGTAGCCATCATCCGCTAT--GAGCAGCATGGCTTCGCC 58
QY 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
DB 59 TTTTCAGAGTGGCCAGGTTTATCTGCTCTTCAGAACTCTCCAGCAAGTTCAGTACTT 118
QY 43 GlnAsnLeuThrThrTyrGluLeuValLeuTyrGlnProValThrAlaAspLeuLeu 62
DB 119 CAGATCTTACTACACGATGATGAGTCTCTCTGCGAGCCAGTGCAGCTGAATTCATC 178
QY 63 ValLysLysGlnValHisPheValAsnAlaSerAspValAsnValLysAla 82
DB 179 GAGAGAAAAAGAAAGTCCACTTTTGTGAATGCGTCTGTGATGTCGACAGTTCAAAGCG 238
QY 83 HisLeuAsnValSerGlyLeuProCysSerValLeuAlaAspValGluAspLeuLeu 102
DB 239 CATTTAAATGTGAGCAGAAATCCATTTAACGTTCTGTAGAACAGTGGAGGACCTAAT 298
QY 103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
DB 299 GAACAGCAGACTTTCATGACAGCTGACGCCCGCGCTCGCTTCATCTATGACGAG 358
QY 123 TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
DB 359 TATCACTCGCTAAATGAATCTATTCTCTGATAGAAAGTCATAACTGAACAGCATCCTGAC 418
QY 143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
DB 419 ATGCTCCAGAAATCTACATCGGATCATCATTCGAGAAAGTACCCACTTATGTTTAAAG 478
QY 163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
DB 479 GTCTCAGGAAGAACAAAGAAATCAAAATGCATCTGGATCGACTTGGATCCATGCC 538
QY 183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 197
DB 539 AGAGAATGATTTACCTGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 598
QY 197 ----- 197
DB 599 GGGAAAGAAATCTGTATACACAGACTTCTGAGGACGTTGATTTCTACATCATGCCGTG 658
QY 198 -----AsnArgMetTrpArgLysAsnArg 205
DB 659 ATGACGTGGATGGCTATGACTACAGTGGAAAGAAATCGAATGTGGAGGAGAACCGT 718
QY 206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
DB 719 ---CTGTCAAGAAACAAACCGTGGTGGGCCAGACCTGAACAGAACTTCGCTTCCAAA 775
QY 226 HisTrpCysGluGluGlnValAsnSerSerSerSerGlu-ThrTyrCys-GlyLeu 245
DB 776 CACTGTTGTGAGAAAGGTCGTCACATGCTCTCTGCTGAAACCTACTGTGGGACTTT 835
QY 245 YrProGluSerGluPro 250
DB 245 YrProGluSerGluPro 250
```

836 ATCCTGAGTCTGAGCCG 852

RESULT 6
LOCUS

DEFINITION

602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',

mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc.

cDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LUCM1595 row: c column: 05

High quality sequence stop: 666.

Location/Qualifiers

1. 738

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4733500"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC_76"

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:

SfiI (ggcgcctcgccc); Site 2: SfiI (ggcattatggcc); 5'-and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CAGCGCATATGCGC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGCGCGCGCATATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.85

kb (range 1.0-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH MGC Library."

BASE COUNT 209 a 161 c 153 g 215 t

ORIGIN

Alignment Scores:

Pred. No.: 1.7e-103 Length: 738

Score: 996.50 Matches: 205

Percent Similarity: 72.39% Conservativity: 10

Best Local Similarity: 69.02% Mismatches: 27

Query Match: 52.15% Indels: 55

DB: 10 Gaps: 3

US-09-980-881A-2 (1-360) x BG616456 (1-738)

QY 10 ValProIleValLeuPheCysGluGlnHisValPheAlaPheGlnSerGlyGlnValLeu 29

DB 1 GTACCATGTTCTCTCTCTGTGAGCAGCATGTCTCGGCTTTCAGATGGCCAGTCTTA 60

QY 30 AlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrTyr 49

DB 61 GCTGCTCTCTAGAACCTCTAGGCAAGTCAAGTCTTACAGAACTCTTACTACACATAT 120

QY 50 GluIleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGlnValHis 69

DB 121 GAGATTTGTTCTCTGCGACCGGTAACTGACCTTATTGTGAAGAAAAACAAGTCCAT 180

QY 70 PhePheValAsnAlaSerAspValAspValLysAlaHisLeuAsnValSerGlyIle 89

```

181 TTTTGTAAATGCTGCTGATGTCGCAATGTGAAGCCCAATTTAAATGTGACGGAATT 240
90 ProCysSerValLeuLeuAlaAspValGluAspLeuLeuGlnGlnGlnSerAsnAsp 109
241 CCATGCGAGTGTCTGCTGCGACAGCTGGAAGATCTTTATTCACACAGAGATTTCCACAGC 300
110 ThrValSerProArgAlaSerAlaSerTyTyTyGluGlnTyHisSerLeuAsnGluLe 129
301 ACAGTCAGCCCCCGAGCCCTCCGATCGTATGATGACAGTATCACTCACTAAATGAATC 360
130 TyrSerTrpIleGluPheIleThrGluArgHisProAspMetLeuThrIleHisIle 149
361 TATTTTGTGATGAGATTTATTAATCTAGAGGCATCTGATATCTTACAAAATCCACATT 420
150 GlySerSerPheGluLysTyTyProLeuTyValLeuLysValSerGlyLysGluGlnThr 169
421 GCATCTCTCAATTTGAGAAGTACCCTCTATGTTTAAAGTTCTTGGAAAGCAACGCA 480
170 AlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAla 189
481 GCCAAAATGCGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
190 PheCysLeuTrpPheIleGlyHis--AsnArgMetTrpArgLysAsnArgSerPheTyra 209
541 TTCTGCTGTGTGTTTATGAGGCATTAATCACTCAATCTATGCGATATAGGCGCAATATACC 600
209 LaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysG 229
601 AATCTCTCGAGCGCTGTGGA-----TTTCTATGTT 630
229 LuGluGlyAlaSerSerSerCysSerGluThrTyTyCysGlyLeuTyTyProGluSerG 249
631 ATGCGCGTGTGTTCA-----TGTGGACTG----- 653
249 LuProGluValLysAlaValAlaSerPheLeuArgAsnIleAsnGlnIleLysAlaIat 269
653 ----- 653
269 TyrIleSerMetHisSerTyTySerGlnHisIleValPheProTyTySerTyTyThrArgSerL 289
654 -----GTTATGGACTACTCTATGACCAAGATC 681
289 ySerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
682 GCATCTGTTGGACAAAAGGAAACCGCTCTTTACTAATGCGAACAATT 728

```

RESULT 7

BY704293

LOCUS BY704293 RIKEN full-length enriched, 18-day embryo whole body Mus
DEFINITION musculus cDNA clone 1110032P04 5', mRNA sequence.

ACCESSION BY704293

VERSION BY704293.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE 1 (bases 1 to 1004)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaoka, I., Kiyosawa, H.,
Yagi, K., Tonari, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,
Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,
Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chochia, C., Corbani,
L. B., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,
A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,
Jarvis, E. D., Kim, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,
King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,
P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Mikki,
H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G.,

Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,
Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,
B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou,
M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,
R. D., Tomita, M., Verdano, R., Wagner, L., Wahlestedt, C., Wang, Y.,
Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa,
M., Yang, I., Yang, B., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii,
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,
S., Hashizume, M., Hayashida, K., Hirozane, T., Hori, F., Inotani, K.,
Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,
H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,
Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
Submission

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

Location/Qualifiers

1..1004

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="1110032P04"

/tissue_type="whole body"

/dev_stages="18-day embryo"

/clone_lib="RIKEN full-length enriched, 18-day embryo

whole body"

BASE COUNT 286 a 228 c 228 g 262 t

ORIGIN

Alignment Scores:

Pred. No.: 2,01e-102 Length: 1004

Score: 989.00 Matches: 199

Percent Similarity: 75.69% Conservative: 19

Best Local Similarity: 69.10% Mismatches: 31

Query Match: 51.75% Indels: 42

Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
 Db 380 GAACAGTATCACTCACTAAATGAATCTATTCTTGATAGATTTTATTAACGAGAGCAT 439

Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
 Db 440 CCTGATATGCTTACAAAATCCACATTTGGATCTCTATTGAGAGTACCCACTCTATGTT 499

Qy 161 -LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
 Db 500 TCTAAGAGTTCTGGAAGAAGAACACAGCAGCCAAATGCCATATGATGACTGTGGAA 559

Qy 180 eHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--AsnArg 199
 Db 560 CCATCCAGAGATGGATCTCTCCGCTTTCTGCTGTGGTTCATAGGCCATATAACTCA 619

Qy 200 MetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIleGly 215
 Db 620 ATTCTATGGGATAATAGGCAATATACCAATCTCTGAGGCTGTGGGA 667

RESULT 9
 BG618239
 LOCUS 602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5',
 DEFINITION mRNA sequence.

ACCESSION BG618239
 VERSION BG618239.1 GI:13669610
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 725)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCW1629 row: f column: 17
 High quality sequence stop: 723.

FEATURES

source
 1..725
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4767376"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_76"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
 SfiI (ggcccctcgcc); Site 2: SfiI (ggccattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGSCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGGAGCGCCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."

BASE COUNT 203 a 153 c 149 g 220 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4,85e-101 Length: 725
 Score: 975.00 Matches: 194
 Percent Similarity: 98.48% Conservatives: 0
 Best Local Similarity: 98.48% Mismatches: 3

Query Match: 51.02% Indels: 2
 DB: 10 Gaps: 0

US-09-980-881A-2 (1-360) x BG618239 (1-725)

Qy 1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
 Db 65 ATGAAGCTTTTCAGCCCTTGCACTCTTGTAGCTCTCTCTAGAACCTCTAGCAAGTTCAA 124

Qy 21 PheAlaPheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGln 40
 Db 125 TTCGGTTTCAGAGTGGCCAAAGTTCTAGCTCTCTCTAGAACCTCTAGCAAGTTCAA 184

Qy 41 ValLeuGlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
 Db 185 GTTCTACAGATCTTACTACCAATATGAGAT-GTTCTCTGGCAGCCGGTAACAGCTGAC 243

Qy 61 LeuIleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
 Db 244 CTTAT-GTGAAGAAAAAACAGTCCATTTTTTTTAAATGCATCTGATGTGCAATGTG 302

Qy 81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
 Db 303 AAACCCATTAAATGTGAGCGGAATTCATGCACTGCTCTGCTGGCAGATGTGAAGAT 362

Qy 101 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
 Db 363 CTTATTCAACAGCAGATTTCCACGACACAGTACGCCCCGAGCTCCGATCTGACTAT 422

Qy 121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
 Db 423 GAACAGTATCACTCACTAAATGAATCTATTCTTTGGATAGAAATTTATAAATGAGAGGAT 482

Qy 141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
 Db 483 CTTGATATGCTTACAAAATCCACATTTGGATCTCTCTTTGAGAAAGTACCCACTCTATGTT 542

Qy 161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
 Db 543 TTTAAAGGTTTCTGAAAAGAACACAGCAGCCAAATGCCATATGATGATGCTGGAATC 602

Qy 181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
 Db 603 CATGCCAGAGATGGATCTCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCAT 653

RESULT 10
 BG569281
 LOCUS 602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5',
 DEFINITION mRNA sequence.

ACCESSION BG569281
 VERSION BG569281.1 GI:13576934
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 852)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCW1580 row: b column: 19
 High quality sequence stop: 751.

FEATURES
 Location/Qualifiers

```
source
1. .852
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:472354"
/lab host="DH10B (T1 phage-resistant)"
/clone lib="NIH MGC 76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccctcgcc); Site 2: SfiI (ggccattatggcc); 5'-and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence:
5'-ATTAGAGCGCGCGCGGCACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT 251 a 177 c 187 g 237 t
ORIGIN

Alignment Scores:
Pred. No.: 8,788-100 Length: 852
Score: 965.00 Matches: 194
Percent Similarity: 75.38% Conservative: 2
Best Local Similarity: 74.62% Mismatches: 7
Query Match: 50.50% Indels: 58
DB: 10 Gaps: 2

US-09-980-881A-2 (1-360) x BG569281 (1-852)
QY 157 ProLeuTyrValLeuLysValSerGlyLysGluInThrAlaLysAlaIleTrpIle 176
DB 2 CCACCTATGTTTAAAGGTTTCTGGAAAGAAAGAACGAGCGCAAAATGCCATGATTT 61
QY 177 AspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGly 196
DB 62 GACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTTTGTTGTTTATAGGC 121
QY 197 His----- 197
DB 122 CATATACTCAATTCATGGGATAATAGGCAATATACCAATCTCTCAGGCTTGTGGAT 181
QY 198 -----AsnArg 199
DB 182 TTCTATGTTATCCAGTGTTAAATGTGGATGTTATGACTACTCATGGAAGAAAGATCGA 241
QY 200 MetTrp-ArgLysAsnArgSerPheTyrAlaHisHisCysIleGlyThrAspLeuAs 219
DB 242 ATGTGGACGAAAGAACCGTTCTTTCTATCGGAACAATCATTCATCGGAACAGACCTGAA 301
QY 219 nSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerG1 239
DB 302 TAGGAACCTTGTCTCCAAACACTGTGTGAGGAAGGTGCATCCAGTTCTCATGCTCGGA 361
QY 239 uThrTyrCysGlyLeuTyrProGluSerProGluValLysAlaValAlaSerPheLe 259
DB 362 AACCTACTGTGGACTTATTCCTAGTCAGAACAGAGTGAAGCAGTGGCTAGTTTCTT 421
QY 259 uArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisI 279
DB 422 GAGAGAAATATCAACACAGATTAAAGCATACATCAGCATGCAATTCATCTCCAGCATAT 481
QY 279 eValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVa 299
DB 482 AGTGTTTCCATATTCCTATACAGCAAGTAAAGCAAGAACCAATGATGAGGAACCTGCTCTAGT 541
QY 299 lAlaSerGluAlaValAtgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisG1 319
DB 542 ACCCGATGAACGAGTTCGTGCTATTGAGAAATATTAGTAAATAATACAGGTATACATGC 601
QY 319 yHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyr-AspL 339
DB 602 GCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGACGATTGGATCTATGCATT 661

339 euGlyIleLysTyrSerPhe----- 345
662 TGGGCATCAGATATTCGTTTACAATTGAACCTTCAGATACCGGCACATAACGATTTCTT 721
346 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeu 359
722 GCTGCCGCGGCGTTTACATCAA-CCACTGTAGAGAAGCTTTTCCCGCTGTCTCTA 776

BI332659 1022 bp mRNA linear EST 30-JUL-2001
602984408F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5137369 5',
mRNA sequence.
BI332659
ACCESSION BI332659
VERSION BI332659.1 GI:15017316
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 1022)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11337 row: 0 column: 02
High quality sequence stop: 841.
FEATURES
Location/Qualifiers
1..1022
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FVB/N"
/clone="IMAGE:5137369"
/lab host="DH10B (T1 phage-resistant)"
/clone lib="NCI_CGAP_L19"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 303 a 233 c 225 g 261 t
ORIGIN

Alignment Scores:
Pred. No.: 8,498-99 Length: 1022
Score: 957.50 Matches: 209
Percent Similarity: 72.81% Conservative: 24
Best Local Similarity: 65.31% Mismatches: 44
Query Match: 50.10% Indels: 46
DB: 12 Gaps: 3

US-09-980-881A-2 (1-360) x BI332659 (1-1022)
QY 3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
DB 3 CTTTCATGCGCTTGGAAATCCTGTAGCCATCATCTCTAT--GAGCAGCATGGCTTCGCC 59
QY 23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
DB 60 TTTTCAGAGTGGCCAGGTTTATCTGCTCTTCCAGAACCTCCAGCAAGTTCACACTATT 119
QY 43 GlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
DB 120 CAGAACTTACTACAAACGATATGAGTCTGTTCTCTGGCAGCCAGTCAGAGCTGAATTCATC 179
```



```

Db      547 GATCTCTCTGCTTTCTGTTGGTGGGGTCATAGGCCAT 583
RESULT 13
BG618813
LOCUS   602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767885 5',
DEFINITION mRNA sequence.
ACCESSION BG618813
VERSION   BG618813.1 GI:13670184
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 750)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs@mail.nih.gov
          Tissue Procurement: CLONETECH Laboratories, Inc.
          cDNA Library Preparation: CLONETECH Laboratories, Inc.
          Mammalian Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/ILNL at:
          http://image.llnl.gov
          Plate: LLCW1630 row: k column: 22
          High quality sequence stop: 622.
          Location/Qualifiers
            1..750
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:4767885"
              /lab_host="DH10B (T1 phage-resistant)"
              /clone_lib="NIH MGC 76"
              /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
              SfiI (ggccctatggcc); Site 2: SfiI (ggccctatggcc); 5' and
              3' adaptors were used in cloning as follows: 5' adaptor
              sequence: 5'-CACGGCCATATGCCC-3' and 3' adaptor sequence:
              5'-ATTCTAGAGCGCGCGCGGCATG-dt(30)BN-3' (where B = A,
              C, or G and N = A, C, G, or T). Average insert size 1.85
              kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
              by PCR. This library was enriched for full-length clones
              and was constructed by Clontech Laboratories (Palo Alto,
              CA). Note: This is a NIH MGC Library."
BASE COUNT 219 a 146 c 171 g 213 t 1 others
ORIGIN
source
Alignment Scores:
Pred. No.: 1.44e-95 Length: 750
Score: 927.50 Matches: 190
Percent Similarity: 95.02% Conservative: 1
Best Local Similarity: 94.53% Mismatches: 6
Query Match: 48.53% Indels: 5
DB: 10 Gaps: 0

US-09-980-881A-2 (1-360) x BG618813 (1-750)
QY      1 MetLysLeuCyseSerLeuAlaValLeuValProLleValLeuPheCysGluGlnHisVal 20
Db      11 ATGAGACTTTGAGGCTTGACAGCTCTTGACCATGTCTCTCTGAGCAGCATGTC 70
QY      21 PheAlaPheGlnSerGlyGlnValLeuAlaLaLeuProArgThrSerArgGlnValGln 40
Db      71 TTTCGGTTTCAGAGTGGCCAAAGTCTTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAA 130
QY      41 ValLeuGlnAsnLeuThrThrThrTyrGluLleValLeuTrpGlnProValThrAlaasp 60
Db      131 GTTCTACAGAAATCTTACTACAAATATGAGATTGTCTCTGCGACCGGTACAGCTGAC 190
QY      61 LeuLleValLysLysGlnValHisPhePheValAsnAlaSerAspValAspVal 80
Db      191 CTTATTGTGAAGAAAAACAGTCCATTTTGTAAATGATCATCTGATGTGCAATGTG 250
QY      81 LysAlaHisLeuAsnValSerGlyLeProCyseSerValLeuLeuAlaAspValGluasp 100
Db      251 AAAGCCCATTTAAATGTGAGCGGAATTCATGCAAGTGTCTTGTGGCAGAGCTGAAGAT 310
QY      101 LeuLleGlnGlnGlnLleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
Db      311 CTTATTCAACAGCAGATTTCCACGACACAGTCCAGCCCGAGCTCCGCACTGCTACTAT 370
QY      121 GluGlnTyrHisSerLeuAsnGluLleTyrSerTrpLleGluPheLleThrGluArgHis 140
Db      371 GAACAGTATCACTCACTAAATGAATCTATTCTTGGATAGATTTATAACTGAGAGGCAT 430
QY      141 ProAspMet-LeuThrLysLleHisLleGly-SerSerPheGluLysTyrProLeu-Tyr 159
Db      431 CCTGATATGCTTAACAAAAATCCACATTTGGCATCTCTCTGCTTCTGCTTGTGTC-ATAGGCCAT 490
QY      160 ValLeuLysValSer-GlyLysGluGlnThrAlaLysAsnAlaLleTrpLleAspCysG1 179
Db      491 GTTTTAAAGGTCTCTGGGAAAAGAACACAGCAGGCAAAAATGCCATATGGATTGACTGTGG 550
QY      179 YLleHisAlaArgGluTrpLleSerProAlaPheCysLeuTrpPheLleGlyHis 197
Db      551 AATCATGCCAGAGATGATCTCTCTGCTTCTGCTTGTGTC-ATAGGCCAT 604

RESULT 14
LOCUS   CB157619
DEFINITION K-EST0216807 L17N670205n1 Homo sapiens cDNA clone
ACCESSION CB157619
VERSION   CB157619.1 GI:28142751
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 541)
AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
          Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
          Kim,Y.S.
TITLE     21C Frontier Korean EST Project 2001
JOURNAL   Unpublished
COMMENT   Contact: Kim YS
          Genome Research Center
          Korea Research Institute of Bioscience & Biotechnology
          52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
          Tel: +82-42-860-4470
          Fax: +82-42-860-4409
          Email: yongsung@mail.kribb.re.kr
          Plate: 12 row: A column: 11
          High quality sequence stop: 541.
          Location/Qualifiers
            1..541
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="L17N670205n1-12-A11"
              /sex="F"
              /lab_host="Top10P"
              /clone_lib="L17N670205n1"
              /note="Organ: Liver; Vector: pT73-Pac; Site 1: EcORI;
              Site 2: NotI; The library was contributed by the Soares
              laboratory and it was constructed as described by Bonaldo,
              M.F., Lennon, G. and Soares, M.B. (1996). Genome Research
              6(9): 791-806. RNA was prepared from harvested cell
              culture."
BASE COUNT 156 a 121 c 107 g 157 t
ORIGIN
Alignment Scores:

```



```
Pred. No.: 4,79e-95 Length: 541
Score: 921.00 Matches: 178
Percent Similarity: 99.44% Conservative: 0
Best Local Similarity: 99.44% Mismatches: 1
Query Match: 48.19% Indels: 0
DB: 14 Gaps: 0

US-09-980-881A-2 (1-360) x CB157619 (1-541)

Qy 3 LeuCySerLeuAlaValLeuValProLeuValLeuPheCysGluGlnHisValPheAla 22
Db 3 CTTTCGACGCTTGCACTGCTTGATCCCATTTCTCTCTCTGAGCAGCATGCTCTTCGGG 62
Qy 23 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 42
Db 63 TTTTCAGAGTGCCAAAGTTCTAGCTCTCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 122
Qy 43 GlnAenLeuThrThrThrGluLeuValLeuValLeuValLeuValLeuValLeuValLeu 62
Db 123 CAGAACTTTACTACAAACATATGAGATTGTTCTCTGCGACCGGTAAACAGCTGACCTTATT 182
Qy 63 VallysLysGlnValHisPheValAenAlaSerAspValAspValAspValLysAla 82
Db 183 GTGAAGAAAAACAAAGTCCATTTTGTAAATGATCTGATGTCGACAAATGTGAAGCC 242
Qy 83 HisLeuAenValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeuLeu 102
Db 243 CATTTAAATGTGAGCGGAATTCATGACAGTGTCTTGTCTGCGAGATGTGGAAGATCTTATT 302
Qy 103 GlnGlnGlnLeuSerAspThrValSerProArgAlaSerAlaSerTyrrTyrrGluGln 122
Db 303 CAACAGCAGATTTCACAGACACATGACGCCCCGAGCCTCGCATCTGATCTATGAAACAG 362
Qy 123 TyrHisSerLeuAenGluLeuTyrrSerTrpLeuGluPheLeuThrGluArgHisProAsp 142
Db 363 TATCACTCACTAAATGAATCTATTCTTGGATAGATTATTAACATGAGAGCATCTTGAT 422
Qy 143 MetLeuThrLysLeuHisLeuGlySerSerPheGluLysTyrrProLeuTyrrValLeuLys 162
Db 423 ATGCTTACAAAATCCACATTGGATCTCTCAATTGAGAAAGTACCCACTCTATGTTTAAAG 482
Qy 163 ValSerGlyLysGluGlnThrAlaLysAsnAlaLeuTrpLeuAspGlyLeuHis 181
Db 483 GTTCTGGAAAGAAACAGACGCCAAAAATGCCATATGATGATTGCTGGAAATCCAT 539

RESULT 15
BG565100 747 bp mRNA linear EST 10-APR-2001
LOCUS 602583720F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711348 5',
DEFINITION mRNA sequence.
ACCESSION BG565100
VERSION BG565100.1 GI:13572753
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CN1551 row: h column: 05
High quality sequence stop: 714.
Location/Qualifiers
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:00:56 ; Search time 58.9207 Seconds
(without alignments)
910.539 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSQVLAALPRTSRQVQL.....IKYSFTSNPPVKEKLLPLSLK 338

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1798	100.0	338	21	Human brain carbox
2	1785	99.3	360	21	Human brain carbox
3	1722	95.8	386	24	Human protease pro
4	1717	95.5	423	17	Human protease pro
5	1699.5	94.5	423	18	Human plasma carbo
6	1699.5	94.5	423	18	Human plasma carbo
7	1699.5	94.5	423	20	Human plasma carbo
8	1694.5	94.2	423	14	Human plasma carbo
9	1338	74.4	246	21	Human brain carbox

10	816.5	45.4	211	21	AA58129	Lung cancer associ
11	646.5	36.0	404	16	AA75131	Porcine Tyr-His-Me
12	637.5	35.5	415	17	AAW06172	Human pancreatic c
13	637.5	35.5	415	19	AAW74476	Human pancreatic c
14	637.5	35.5	417	24	ABB99688	Amino acid sequenc
15	637.5	35.5	424	17	AAW06175	ProHCPB with PeIB
16	636.5	35.4	324	22	AAW06175	Bothops jararaca c
17	622	34.6	417	23	AAU84325	Protein CPA3 diffe
18	622	34.6	667	24	ABP97414	Anti-human seminal
19	619.5	34.5	716	18	AAW41414	PreproHCPB-linker-
20	612	34.0	434	24	ABB99687	Human carboxypepti
21	612	34.0	437	22	AAE00502	Amino acid sequenc
22	612	34.0	437	23	AAE19175	Human carboxypepti
23	610	33.9	307	17	AAW00602	Rat carboxypeptida
24	582	32.4	306	16	AAW75132	Porcine carboxypep
25	582	32.4	306	22	AAU04477	Porcine carboxypep
26	569	31.6	354	22	AAW06173	Human secreted met
27	569	31.6	354	22	AAW06173	Human secreted met
28	569	31.6	362	22	AAW74682	Human protease and
29	569	31.6	374	22	AAW06173	Human secreted met
30	569	31.6	374	22	AAW06173	Human secreted met
31	569	31.6	374	23	AAU82703	Amino acid sequenc
32	569	31.6	374	24	ABJ37888	NOVX protein seque
33	569	31.6	374	24	ABG74658	Human zinc-binding
34	569	31.6	444	23	AAE24240	Human 23566 (carbo
35	567	31.5	329	17	AAW06174	Mature HCPB with P
36	567	31.5	349	17	AAW06173	Modified HCPB (D25
37	560	31.1	349	17	AAW06181	Carboxypeptidase B
38	560	31.1	349	18	AAW13749	Carboxypeptidase B
39	559	31.1	349	17	AAW06182	Modified HCPB (D25
40	559	31.1	349	18	AAW13750	Carboxypeptidase B
41	556	30.9	349	18	AAW13763	Carboxypeptidase B
42	554	30.8	349	18	AAW13752	Carboxypeptidase B
43	554	30.8	349	18	AAW13758	Carboxypeptidase B
44	554	30.8	399	22	AAE01664	Novel human protea
45	553	30.8	349	18	AAW13751	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAB11458
ID AAB11458 standard; Protein; 338 AA.
XX
AC AAB11458;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 3.
XX

KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; neurotropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.

OS Homo sapiens.

PN WO2000066717-A1.

PD 09-NOV-2000.

PF 01-MAY-2000; 2000WO-JP02878.

PR 30-APR-1999; 99JP-0125169.

PA (MATS/) MATSUMOTO A.

PI Matsumoto A;

XX WPI; 2000-687534/67.

CC protein therapy) and diagnostic compositions. The peptides are also
 CC useful for raising antibodies or eliciting an immune response (vaccine),
 CC as a reagent (including the labeled reagent) in assays designed to
 CC quantitatively determine levels of the protein (or its binding partner or
 CC ligand) in biological fluids, or as markers for tissues in which the
 CC corresponding protein is preferentially expressed and in methods for
 CC identifying a modulator of the peptide or an agent that binds to the
 CC peptide. The agents identified are useful for treating protease-related
 CC conditions that are specific for the subfamily of proteases that the
 CC peptide belongs to, particularly in cells and tissues that express the
 CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
 CC disorders. The modulator of the peptide is also useful for treating a
 CC disorder characterised by an absence of, inappropriate or unwanted
 CC expression of the protein. The sequence presented is the human protease
 CC protein, variant #1, the gene for which is located on chromosome 13.

XX
 SQ Sequence 386 AA;
 Query Match 95.5%; Score 1717; DB 24; Length 386;
 Best Local Similarity 99.4%; Pred. No. 4.5e-171;
 Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTDLIYKKQVHFFVNASDVNDVKA 60
 DB 23 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTDLIYKKQVHFFVNASDVNDVKA 82
 QY 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD 120
 DB 83 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD 142
 QY 121 MLTKIHGSSFEKYPYLYLVKVGKQTAQNAIWDGCIHAREWISPAFCFLWFIHNRWVR 180
 DB 143 MLTKIHGSSFEKYPYLYLVKVGKQTAQNAIWDGCIHAREWISPAFCFLWFIHNRWVR 202
 QY 181 KNRSFYANNHCIGTDLNRNFASKHCEGASSSCSETYCGLYPSEPEVKAVASFLRN 240
 DB 203 KNRSFYANNHCIGTDLNRNFASKHCEGASSSCSETYCGLYPSEPEVKAVASFLRN 262
 QY 241 INQIKAYISMHSYQSHIYPPYSTRKSKDHLESLVASEAVRAIEKTSKNTRYTHGHGS 300
 DB 263 INQIKAYISMHSYQSHIYPPYSTRKSKDHLESLVASEAVRAIEKTSKNTRYTHGHGS 322
 QY 301 ETLYLAPGGDDWIYDLGIKYSFT 324
 DB 323 ETLYLAPGGDDWIYDLGIKYSFT 346

RESULT 5

AAR90293

ID AAR90293 standard; Protein; 423 AA.

XX AAR90293;

XX 25-MAR-2003 (updated)

DT 12-APR-1996 (first entry)

XX Human plasma carboxypeptidase B.

XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
 KW purification; plasminogen; affinity column.

XX Homo sapiens.
 XX

Key Location/Qualifiers
 FT Active-site 181

FT /note= "Residue suspected of being involved in
 FT catalytic activity of enzyme."

FT Binding-site 183

FT /note= "Residue suspected of being part of the
 FT binding site of enzyme."

FT Binding-site 256..257

FT /note= "Residues suspected of being part of the
 FT binding site of enzyme."

FT Active-site 310
 FT /note= "Residue suspected of being involved in
 FT catalytic activity of enzyme."
 FT Binding-site 312
 FT /note= "Residue suspected of being part of the
 FT binding site of enzyme."
 FT Active-site 369
 FT /note= "Residue suspected of being involved in
 FT catalytic activity of enzyme."
 FT Binding-site 370..371
 FT /note= "Residues suspected of being part of the
 FT binding site of enzyme."
 FT Binding-site 394
 FT /note= "Residue suspected of being part of the
 FT binding site of enzyme."

XX US5474901-A.

XX 12-DEC-1995.

XX 19-JUL-1994; 94US-0277540.

XX 01-FEB-1991; 91US-0649591.

XX 14-OCT-1992; 92US-0959944.

XX 15-DEC-1993; 93US-0167727.

XX 19-JUL-1994; 94US-0277540.

XX (GETH) GENENTECH INC.

XX Drayna DT, Eaton DL;

XX WPI, 1996-039508/04.

XX N-PSDB; AAT11671.

XX Antibody to human plasma carboxypeptidase B - useful for detecting
 PT and purifying hPCPB for use in treating clotting disorders e.g.
 PT haemophilia A

XX Disclosure; Figure 4; 40pp; English.

XX An antibody which specifically binds human plasma carboxypeptidase B
 CC (hPCPB) and does not cross react with other carboxypeptidases is
 CC useful for the detection of hPCPB in vitro. The antibody is also
 CC used for purifying hPCPB from a sample. Purification comprises
 CC passing a sample thought to contain hPCPB over either a column to
 CC which antibody has been bound, or a plasminogen affinity column,
 CC eluting the column and then recovering the fraction containing the
 CC hPCPB.

XX (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 423 AA;

SQ

Query Match 94.5%; Score 1699.5; DB 17; Length 423;
 Best Local Similarity 89.8%; Pred. No. 3.6e-169;
 Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTDLIYKKQVHFFVNASDVNDVKA 60
 DB 23 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTDLIYKKQVHFFVNASDVNDVKA 82
 QY 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD 120
 DB 83 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD 142
 QY 121 MLTKIHGSSFEKYPYLYLVKVGKQTAQNAIWDGCIHAREWISPAFCFLWFIHNRWVR 175
 DB 143 MLTKIHGSSFEKYPYLYLVKVGKQTAQNAIWDGCIHAREWISPAFCFLWFIHNRWVR 202
 QY 176 -----NNWRKRSFYANNHCIGTDLNRNFASK 203
 DB 203 GIIGQYTNLLRLVDYFVMPVNVVDGYDSWKKNRMRKRSFYANNHCIGTDLNRNFASK 262
 QY 204 HWCEBAGASSSCSETYCGLYPSEPEVKAVASFLRNINQIKAYISMHSYQSHIYPPY 263


```
FT Protein 23..423
FT /note= "mature PCPBthr147"
FN WO9855645-A1.
XX 10-DEC-1998.
XX 02-JUN-1998; 98WO-EP03244.
XX 03-JUN-1997; 97US-0869057.
XX (SCHD ) SCHERING AG.
XX Morser MJ, Nagashima M;
XX WPI; 1999-045800/04.
XX N-PSDB; AAV74302.
XX Detecting new polymorphism of human plasma carboxypeptidase B -
XX comprises Alanine or Threonine at position 147 of protein by DNA or
XX protein analysis, useful to detect risk of thrombotic disease in
XX humans
XX
XX Disclosure; Page 25-26; 35pp; English.
XX
XX This sequence is the human plasma carboxypeptidase B (PCPB) mutant
XX hPCPBthr147. The invention relates to a method for determining the
XX presence of DNA or protein polymorphisms of PCPB in human subjects, which
XX comprises obtaining a prepared tissue or blood sample and determining the
XX presence of DNA coding for naturally occurring polymorphs of the protein
XX containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
XX respectively). Determination of the relative distribution of the PCPB
XX polymorphs in a patient's blood by genetic or protein analysis by the
XX methods is useful to determine the risk of thrombotic disease in humans.
XX Such assessments may be made by accumulating information concerning the
XX relative distribution of the different polymorphs within the general
XX population compared with populations known to be at risk and establishing
XX a PCPB polymorph profile for at-risk patients.
XX
XX Sequence 423 AA;
XX
XX Query Match 94.5%; Score 1699.5; DB 20; Length 423;
XX Best Local Similarity 89.8%; Pred. No. 3.6e-169;
XX Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 FQSGQVLAALPRTSRQVQLQLTTTTEIYLWQPTADLIIVKKQVHFFVNASDVNDVKA 60
Db 23 FQSGQVLAALPRTSRQVQLQLTTTTEIYLWQPTADLIIVKKQVHFFVNASDVNDVKA 82
QY 61 HLNVSIGPCSVLLADVEDLIQQIISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD 120
Db 83 HLNVSIGPCSVLLADVEDLIQQIISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD 142
QY 121 MLTKIHGSSFKYPLVYLKVSQKQTAQNAIWDICGIIHAREWISPAFLMFIH----- 175
Db 143 MLTKIHGSSFKYPLVYLKVSQKQTAQNAIWDICGIIHAREWISPAFLMFIHITQFY 202
QY 176 -----NRWRKNSFYANNHCIGTDLNRFASK 203
Db 203 GIIGQYTNLLRLVDFVMPVNVVDGYDSWKKNRWRKNSFYANNHCIGTDLNRFASK 262
QY 204 HWCESGASSSSCSEYCYGLYPESEPEVKAVASFLRNINQIKAYISMHSYSHIYFPYSY 263
Db 263 HWCESGASSSSCSEYCYGLYPESEPEVKAVASFLRNINQIKAYISMHSYSHIYFPYSY 322
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSETLYLAPGGDDDIYDLGIKYSF 323
Db 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGSETLYLAPGGDDDIYDLGIKYSF 382
QY 324 T 324
Db 383 T 383
```

```
RESULT 8
AAR36273
ID AAR36273 standard; Protein; 423 AA.
XX AC AAR36273;
XX DT 24-AUG-1993 (first entry)
XX DE Human plasma carboxypeptidase B.
XX KW PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX OS Homo sapiens.
XX FH Key
XX FT Peptide
XX FT 1..22
XX FT /label= signal_peptide
XX FT Cleavage-site 114
XX FT /note= "putative site of cleavage by trypsin to
XX FT activate PCPB as a carboxypeptidase"
XX FT Active-site 181
XX FT /note= "expected to be involved in catalytic
XX FT activity"
XX FT Binding-site 183
XX FT /note= "expected to be involved in substrate
XX FT binding"
XX FT Active-site 184
XX FT /note= "expected to be involved in catalytic
XX FT activity"
XX FT Binding-site 256..257
XX FT /note= "expected to be involved in substrate
XX FT binding"
XX FT Active-site 310
XX FT /note= "expected to be involved in catalytic
XX FT activity"
XX FT Binding-site 312
XX FT /note= "expected to be involved in substrate
XX FT binding"
XX FT Active-site 369
XX FT /note= "expected to be involved in catalytic
XX FT activity"
XX FT Binding-site 370..371
XX FT /note= "expected to be involved in substrate
XX FT binding; Asp 370 is the residue which
XX FT determines specificity of the PCPB as
XX FT a carboxypeptidase B"
XX FT Binding-site 394
XX FT /note= "expected to be involved in substrate
XX FT binding"
XX US206161-A.
XX 27-APR-1993.
XX 01-FEB-1991; 91US-0649591.
XX 01-FEB-1991; 91US-0649591.
XX (GETH ) GENENTECH INC.
XX Drayna DT, Eaton DL;
XX WPI; 1993-151724/18.
XX N-PSDB; AAQ41001.
XX New human plasma carboxypeptidase B - used as haemostatic
XX regulator for clotting blood, partic. for treating blood clotting
XX disorders, e.g. haemophilia
XX Claim 2; Fig 4; 40pp; English.
XX This amino acid sequence was deduced from a clone isolated from a
```

CC human liver cDNA library. There is about 40% sequence identity
CC between the deduced preproPCPB amino acid sequence and pro-human
CC mast cell carboxypeptidase A and between preproPCPB and prepro-rat
CC carboxypeptidase B. Human PCPB has the same substrate binding sites
CC as, and shares 6 cysteine residues which form 3 intramolecular
CC disulphide bonds with, bovine and rat carboxypeptidase B. The
CC presence of the same amino acid (Asp) at position 348 of the mature
CC protein) at the region in carboxypeptidases that determines
CC substrate specificity, suggests that PCPB represents a plasma-
CC derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC tPA of plasminogen to plasmin in the presence of fibrinogen.
XX
SQ Sequence 423 AA;

Query Match 94.2%; Score 1694.5; DB 14; Length 423;
Best Local Similarity 89.5%; Pred. No. 1.2e-168;
Matches 323; Conservative 0; Mismatches 1; Indels 37; Gaps 1;

QY 1 FOSGQVLAALPRTSRQVQLNLTITTYEIVLWQPVTDLIKKKQVHFVNASDVNVKA 60
DB 23 FOSGQVLAALPRTSRQVQLNLTITTYEIVLWQPVTDLIKKKQVHFVNASDVNVKA 82

QY 61 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYVEQYHSLNEIYSWIEFITERHPD 120
DB 83 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYVEQYHSLNEIYSWIEFITERHPD 142

QY 121 MLTKIHGSSFEKYPYLVKVGSKQQTAKNAIWIDCGIHAREWISPAFCLEWFIH- 175
DB 143 MLTKIHGSSFEKYPYLVKVGSKQQTAKNAIWIDCGIHAREWISPAFCLEWFIHITQFY 202

QY 176 -----NMWRKNSFYANNHCIGTDLNRNFASK 203
DB 203 GIIGQYTNLLRLVDFYVNVVDGYDYSWKKNRMKNSFYANNHCIGTDLNRNFASK 262

QY 204 HWCEGASSSCSEYCYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYQHVFPYSY 263
DB 263 HWCEGASSSCSEYCYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYQHVFPYSY 322

QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSF 323
DB 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSF 382

QY 324 T 324
DB 383 T 383

RESULT 9
AAB11459
ID AAB11459 standard; Protein; 246 AA.
XX
AC AAB11459;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 4.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
FN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX

PA (MATS/) MATSUMOTO A.
XX Matsumoto A;
XX
XX WPI; 2000-687534/67.
XX
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
XX screening agents for the treatment of Alzheimer's and other brain
XX disorders -
XX
XX Disclosure; Page 74-75; 84pp; Japanese.
XX
XX This invention describes a novel protein with peptidase activity
XX against brain beta-amyloid precursor protein which has been isolated from
XX human hippocampus and which has cerebroprotective, antialzheimers,
XX nootropic, neuroprotective and hemostatic activity and which can be used
XX as a vaccine or for gene therapy. The protein, and compounds identified
XX by screening as promoters or inhibitors of its activity, are used to
XX regulate beta-amyloid accumulation in the brain and treat or prevent
XX diseases in which this occurs, such as Alzheimer's, senile dementia,
XX inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 246 AA;

Query Match 74.4%; Score 1338; DB 21; Length 246;
Best Local Similarity 100.0%; Pred. No. 1.3e-131;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 ASASYVQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSKQQTAKNAI 152
DB 1 ASASYVQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSKQQTAKNAI 60

QY 153 WIDCGIHAREWISPAFCLEWFIHGHGSETLYLAPGGDDWIYDLGIKYSFNSPPVEKL 212
DB 61 WIDCGIHAREWISPAFCLEWFIHGHGSETLYLAPGGDDWIYDLGIKYSFNSPPVEKL 120

QY 213 SSCSETYCYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYQHVFPYSYTRSKSKDHE 272
DB 121 SSCSETYCYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYQHVFPYSYTRSKSKDHE 180

QY 273 ELSLVAASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFNSPPVEKL 332
DB 181 ELSLVAASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFNSPPVEKL 240

QY 333 LPLSLK 338
DB 241 LPLSLK 246

RESULT 10
AAB58129
ID AAB58129 standard; Protein; 211 AA.
XX
AC AAB58129;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polypeptide sequence SEQ ID 467.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antineoplastic; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
OS Homo sapiens.
XX
FN WO200055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX

PR	12-MAR-1999;	99US-0124270.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	(ROSE/) ROSEN C A.	
XX		
XX	Ruben SM;	
PI		
XX		
DR	WPI; 2000-587514/55.	
PS	N-PSDB; AAF18005.	
XX		
XX		
PT	Lung cancer associated gene sequences, referred to as lung cancer	
PT	antigens, useful for treatment, prevention, and diagnosis of disorders	
PT	such as lung cancer -	
XX		
XX	Claim 11; Page 944-945; 1425pp; English.	
XX		
CC	Poly-nucleotide sequences AAF17982 - AAF18424 encode human lung cancer	
CC	associated proteins represented in AAB58106 - AAB58548. Lung cancer	
CC	associated proteins and poly-nucleotide sequences, their agonists, and	
CC	antagonists may have neuroprotective; cytostatic; cardioactive;	
CC	immunomodulatory; muscular active general; vulnertary; Gastrointestinal	
CC	general; nephrotropic; anti-infective; gynecological; or antibacterial	
CC	activity. The invention also includes antibodies specific for the	
CC	protein or poly-nucleotide sequences. The lung cancer associated	
CC	poly-nucleotide sequences may be used for detection of lung cancer,	
CC	chromosome identification, as chromosome markers, and for numerous other	
CC	diagnostic or research purposes. The proteins may be used to treat	
CC	disorders such as neural, immune, muscular, reproductive,	
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative	
CC	disorders. The proteins may also be used in the treatment of wounds and	
CC	infectious diseases. Poly-nucleotide sequences AAF18425 - AAF18433 and	
CC	peptide AAB58549 are used in the course of the invention for the	
CC	identification and characterisation of the polynucleotide and protein	
CC	sequences.	
XX		
XX		
SQ	Sequence 211 AA;	
	Query Match 45.4%; Score 816.5; DB 21; Length 211;	
	Best Local Similarity 90.6%; Pred. No. 5.7e-77;	
	Matches 154; Conservative 1; Mismatches 6; Indels 9; Gaps 2	
OY	163 WISPAFCFLMFI-----GHNRMWRKNRSPYANNHCIGTDLNRNPFASKHWCEGASSSS 214	
Db	3 WIS-MLCRWLMMVMNYMSWKNNRMWRKNRSPYANNHCIGTDLNRNPFASKHWCEGASSSS 61	
OY	215 CSETYCGLYPESEPEVKAVASFLERRNIQIKAYISMHSYSQHIVFPYTSKSKDHEEL 274	
Db	62 CSETYCGLYPESEPEVKAVASFLERRNIQIKAYISMHSYSQHIVFPYTSKSKDHEEL 121	
OY	275 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 324	
Db	122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 171	
RESULT 11 .		
AAR75131		
ID	AAR75131 standard; Protein; 404 AA.	
XX		
AC	AAR75131;	
XX		
DT	13-MAR-1996 (first entry)	
XX		
DE	Porcine Tyr-His-Met Procarboxypeptidase B.	
XX		
KW	Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;	
KW	human serum albumin; premating factor alpha; mating factor alpha;	
XX	pOCBB.	
OS	Sus scrofa.	
XX		
PN	WO9514096-A1.	
XX		
PD	26-MAY-1995.	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:21:46 ; Search time 20.8163 Seconds
(without alignments)
687.014 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQL.....IKYSFTSNPPVKLLPLSLK 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1717	95.5	386	4	US-09-813-133A-2
2	1699.5	94.5	423	1	US-07-649-591B-3
3	1699.5	94.5	423	1	US-08-277-540-3
4	1699.5	94.5	423	1	US-08-430-787A-3
5	1699.5	94.5	423	2	US-08-869-057-2
6	1688.5	93.9	423	4	US-09-813-133A-4
7	646.5	36.0	404	1	US-08-696-139-2
8	637.5	35.5	415	2	US-08-860-882A-57
9	637.5	35.5	415	4	US-09-011-769A-39
10	637.5	35.5	424	4	US-09-011-769A-56
11	632	35.2	417	1	US-07-649-591B-7
12	632	35.2	417	1	US-08-277-540-7
13	632	35.2	417	1	US-08-430-787A-7
14	622	34.6	417	1	US-07-649-591B-6
15	622	34.6	417	1	US-08-277-540-6
16	622	34.6	417	3	US-08-430-787A-6
17	619.5	34.5	716	3	US-08-171-945-125
18	612	34.0	437	4	US-09-675-305-10
19	610	33.9	307	2	US-08-782-760-5
20	610	33.9	307	5	PCT-US96-00995-6
21	609.5	33.9	396	1	US-07-649-591B-4
22	609.5	33.9	396	1	US-08-277-540-4
23	609.5	33.9	396	1	US-08-430-787A-4
24	582	32.4	306	1	US-08-696-139-4
25	567	31.5	329	4	US-09-011-769A-51
26	567	31.5	349	4	US-09-011-769A-47
27	560	31.1	349	4	US-09-011-769A-60

28	559	31.1	349	4	US-09-011-769A-64
29	554	30.8	399	4	US-09-710-099-8
30	549.5	30.6	417	1	US-07-649-591B-8
31	549.5	30.6	417	1	US-08-277-540-8
32	549.5	30.6	417	1	US-08-430-787A-8
33	549	30.5	613	3	US-09-171-945-113
34	534.5	29.7	350	4	US-09-675-305-12
35	532.5	29.6	419	1	US-07-649-591B-5
36	532.5	29.6	419	1	US-08-277-540-5
37	532.5	29.6	419	1	US-08-430-787A-5
38	527.5	29.3	436	4	US-09-710-099-6
39	516.5	28.7	417	3	US-08-640-906-4
40	516.5	28.7	417	4	US-09-395-936-4
41	515.5	28.7	417	3	US-08-640-906-18
42	515.5	28.7	417	4	US-09-395-936-18
43	491.5	27.3	419	3	US-08-640-906-2
44	491.5	27.3	419	4	US-09-395-936-2
45	486.5	27.1	419	3	US-08-640-906-17

Sequence 64, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 113, App
Sequence 12, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 6, Appli
Sequence 4, Appli
Sequence 18, Appli
Sequence 2, Appli
Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO01173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Query Match 95.5%; Score 1717; DB 4; Length 386;
Best Local Similarity 99.4%; Pred. No. 1.5e-176;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	FQSGQVLAALPRTSRQVQLNLTFTTVEIWLQPVTDALIVKKQVHFFVNASVDNVKA	60
Db	23	FQSGQVLAALPRTSRQVQLNLTFTTVEIWLQPVTDALIVKKQVHFFVNASVDNVKA	82
Qy	61	HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD	120
Db	83	HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAYEYQVHSLNEIYSWIEFITERHPD	142
Qy	121	MLTKIHIGSSFEXKPYLYLVKVGKEQAKNAIWDCGIHAREWISPAFLCMFIGNRMR	180
Db	143	MLTKIHIGSSFEXKPYLYLVKVGKEQAKNAIWDCGIHAREWISPAFLCMFIGNRMR	202
Qy	181	KNRSFYANNHICIGTDLNRNFASKHWCCEGSSSCSTCYGLYPESEPEVKAVASFLARN	240
Db	203	KNRSFYANNHICIGTDLNRNFASKHWCCEGSSSCSTCYGLYPESEPEVKAVASFLARN	262
Qy	241	INQIKAVISMHSYQSHIVFPYVYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS	300
Db	263	INQIKAVISMHSYQSHIVFPYVYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS	322
Qy	301	ETLYLAPGGGDDMIYDLGIKYSFT	324
Db	323	ETLYLAPGGGDDMIYDLGIKYSFT	346

RESULT 2

US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; US-07-649-591B-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.3e-174;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

Qy 1 FQSQVLAALPRTSRQVQLNLTYYEIVLWQVPTADLIIVKKQVHFFVNASDVNVKA 60
Db 23 FQSQVLAALPRTSRQVQLNLTYYEIVLWQVPTADLIIVKKQVHFFVNASDVNVKA 82

Qy 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQVHSLNEIYSWIEFTTERHPD 120
Db 83 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQVHSLNEIYSWIEFTTERHPD 142

Qy 121 MLTKIHIGSSPEKPYLYLVKVSQKEQTAKNAIWDGCIHAREWISPAFLWFIGH- 175
Db 143 MLTKIHIGSSPEKPYLYLVKVSQKEQTAKNAIWDGCIHAREWISPAFLWFIGHITQFY 202

Qy 176 -----NRMWRKNRSFYANNHCIGTDLNRNFASK 203
Db 203 GIIGQYTNLLRLVDFYVMPVNVVDGYDSWKCKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy 204 HWCBEGASSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSY 263
Db 263 HWCBEGASSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSY 322

Qy 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPCGGDDWYDLGKYSF 323
Db 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPCGGDDWYDLGKYSF 382

Qy 324 T 324
Db 383 T 383

RESULT 3
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-277-540-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.3e-174;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

Qy 1 FQSQVLAALPRTSRQVQLNLTYYEIVLWQVPTADLIIVKKQVHFFVNASDVNVKA 60
Db 23 FQSQVLAALPRTSRQVQLNLTYYEIVLWQVPTADLIIVKKQVHFFVNASDVNVKA 82

Qy 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQVHSLNEIYSWIEFTTERHPD 120
Db 83 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQVHSLNEIYSWIEFTTERHPD 142

Qy 121 MLTKIHIGSSPEKPYLYLVKVSQKEQTAKNAIWDGCIHAREWISPAFLWFIGH- 175
Db 143 MLTKIHIGSSPEKPYLYLVKVSQKEQTAKNAIWDGCIHAREWISPAFLWFIGHITQFY 202

Qy 176 -----NRMWRKNRSFYANNHCIGTDLNRNFASK 203
Db 203 GIIGQYTNLLRLVDFYVMPVNVVDGYDSWKCKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy 204 HWCBEGASSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSY 263
Db 263 HWCBEGASSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSY 322

```

QY      264 TRSKS KOHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYIAPGGDDWIYDLGIKYSF 322
|       |||||
Db      323 TRSKS KOHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYIAPGGDDWIYDLGIKYSF 382
|       |||||
QY      324 T 324
|
Db      383 T 383

RESULT 4
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430.787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-430-787A-3

Query Match          94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.3e-174;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1

QY      1   FSGGVLAALPRTSRQVOVLQNLTITYIVLVQPVTADLIYKKKQVHFVNASDVNVKA 60
|       |||||
Db      23   FSGGVLAALPRTSRQVOVLQNLTITYIVLVQPVTADLIYKKKQVHFVNASDVNVKA 82
|       |||||
QY      61   HLNVSIGIPCSVLLADVLDLIIQQOISNDTVSPRASASYEQVHSLNEIYSWIFITERHPD 120
|       |||||
Db      83   HLNVSIGIPCSVLLADVLDLIIQQOISNDTVSPRASASYEQVHSLNEIYSWIFITERHPD 142
|       |||||
QY      121 MLTKIHGSSPEKYPLYVLKVSGKEQTAKNAIWDCIGHAREWISPAFLWFIGH----- 175
|       |||||

```

```

Db      143  MLTKIHGSSFEKYPLYVLKVSQKEQTAKNAIWDICGHIAREWISPAFLWPIGHIOTFY 203
Qy      176  -----NEMWRKQRSFYANNHCIGTDLNRNFASK 203
Db      203  GIIGQYTNLRLVDFFYVMPVVVDGYDYSWKKNRWRKQRSFYANNHCIGTDLNRNFASK 262
Qy      204  HWCEEGASSSSCSETYCGLYPESPEPKAVASFLURRNINQIKAYISMHSYSQHIVPFYSY 263
Db      263  HWCEEGASSSSCSETYCGLYPESPEPKAVASFLURRNINQIKAYISMHSYSQHIVPFYSY 322
Qy      264  TRSKSKDHELSLVSASAVRAIEKTSKNTRYTHGSGSETLYLAPGGGDDWIYDLGIKYSF 323
Db      323  TRSKSKDHELSLVSASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGGDDWIYDLGIKYSF 382
Qy      324  T 324
Db      383  T 383

RESULT 5
US-08-869-057-2
; Sequence 2, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: washtien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-5411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; TISSUE TYPE: Plasma
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 23..401
US-08-869-057-2

Query Match 94.5%; Score 1699.5; DB 2; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.3e-174;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1

Qy      1  FOSGOVLAALPRTSRQVQVLQNLTTTVEIVLWQPVTADLIVKKQVHFVFNASVDNVKA 60
Db      23  FOSGOVLAALPRTSRQVQVLQNLTTTVEIVLWQPVTADLIVKKQVHFVFNASVDNVKA 82

```

```

Db      143  MLTKIHGSSFEKYPLYVLKVSQKEQTAKNAIWDCGHIHAREWISPAFLWPIGHIOTFY 203
Qy      176  -----NEMWRKQRSFYANNHCIGTDLNRNFASK 203
Db      203  GIIGQYTNLRLVDFFYVMPVVVDGYDYSWKKNRWRKQRSFYANNHCIGTDLNRNFASK 262
Qy      204  HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLURRNINQIKAYISMHSYSQHIVPFYSY 263
Db      263  HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLURRNINQIKAYISMHSYSQHIVPFYSY 322
Qy      264  TRSKSKDHELSLVSASAVRAIEKTSKNTRYTHGSGSETLYLAPCGGDDWIYDLGIKYSF 323
Db      323  TRSKSKDHELSLVSASEAVRAIEKTSKNTRYTHGSGSETLYLAPCGGDDWIYDLGIKYSF 382
Qy      324  T 324
Db      383  T 383

RESULT 5
US-08-869-057-2
; Sequence 2, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: washtien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-5411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; TISSUE TYPE: Plasma
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 23..401
US-08-869-057-2

Query Match          94.5%; Score 1699.5; DB 2; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.3e-174;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1

Qy      1  FOSGOVLAALPRTSRQVQVLQNLTTTVEIVLWQPVTADLIVKKQVHFVFNASVDNVKA 60
Db      23  FOSGOVLAALPRTSRQVQVLQNLTTTVEIVLWQPVTADLIVKKQVHFVFNASVDNVKA 82

```



```
Qy 61 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYVEQYHSLNEIYSWIEFITERHPD 120
Db 83 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYVEQYHSLNEIYSWIEFITERHPD 142
Qy 121 MLTKIHIGSSPEKYPYLVKVGSGKEQTAKNAIWDGCIHAREWISPAFLWFIQH----- 175
Db 143 MLTKIHIGSSPEKYPYLVKVGSGKEQTAKNAIWDGCIHAREWISPAFLWFIQHITQFY 202
Qy 176 -----NMWRKRSFYANNHCIGTDLNRNFASK 203
Db 203 GIIGQYTNLLRLVDFYVMPVNVDPGYDYSWKKNRMWRKRSFYANNHCIGTDLNRNFASK 262
Qy 204 HCWEGASSSSCSETYCGLYPESBPVKAVASFLRRNINQIKAYISMHSYSQHIHVFVPSY 263
Db 263 HCWEGASSSSCSETYCGLYPESBPVKAVASFLRRNINQIKAYISMHSYSQHIHVFVPSY 322
Qy 264 TRSKSKOHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWYDILGIKYSP 323
Db 323 TRSKSKOHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWYDILGIKYSP 382
Qy 324 T 324
Db 383 T 383

RESULT 6
US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weinui et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-4

Query Match 93.9%; Score 1688.5; DB 4; Length 423;
Best Local Similarity 89.2%; Pred. No. 2e-173;
Matches 322; Conservative 0; Mismatches 2; Indels 37; Gaps 1;

Qy 1 FQSGQVLAALPRTSRQVQVLQNLTTTVEIWLQPVTDADLIYKKQVHFFVNASVDVNVKA 60
Db 23 FQSGQVLAALPRTSRQVQVLQNLTTTVEIWLQPVTDADLIYKKQVHFFVNASVDVNVKA 82
Qy 61 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYVEQYHSLNEIYSWIEFITERHPD 120
Db 83 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYVEQYHSLNEIYSWIEFITERHPD 142
Qy 121 MLTKIHIGSSPEKYPYLVKVGSGKEQTAKNAIWDGCIHAREWISPAFLWFIQH----- 175
Db 143 MLTKIHIGSSPEKYPYLVKVGSGKEQTAKNAIWDGCIHAREWISPAFLWFIQHITQFY 202
Qy 176 -----NMWRKRSFYANNHCIGTDLNRNFASK 203
Db 203 GIIGQYTNLLRLVDFYVMPVNVDPGYDYSWKKNRMWRKRSFYANNHCIGTDLNRNFASK 262
Qy 204 HCWEGASSSSCSETYCGLYPESBPVKAVASFLRRNINQIKAYISMHSYSQHIHVFVPSY 263
Db 263 HCWEGASSSSCSETYCGLYPESBPVKAVASFLRRNINQIKAYISMHSYSQHIHVFVPSY 322
Qy 264 TRSKSKOHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWYDILGIKYSP 323
Db 323 TRSKSKOHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWYDILGIKYSP 382
```

```
Qy 324 T 324
Db 383 T 383

RESULT 7
US-08-696-139-2
; Sequence 2, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hershberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/696,139
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-696-139-2

Query Match 36.0%; Score 646.5; DB 1; Length 404;
Best Local Similarity 38.0%; Pred. No. 4.7e-61;
Matches 137; Conservative 60; Mismatches 121; Indels 43; Gaps 6;

Qy 1 FQSGQVLAALPRTSRQVQVLQNLTTTVEIWLQPVTDADLIYKKQVHFFVNASVDVNVKA 60
Db 10 FEQEKVRVNVVEDNDISLLHELASTRQIDFWKPDSTQIKPHSTVDVFRVKAEDILAVED 69
Qy 61 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYVEQYHSLNEIYSWIEFITERHPD 120
Db 70 FLEQNEIQYEVLLNNLSRVLEAQPDSRV---RTTGHSEYKYNWETLEAWTKQVTSNP 126
Qy 121 MLTKIHIGSSPEKYPYLVKVGSGKEQTAKNAIWDGCIHAREWISPAFLWFI----- 173
Db 127 LISRTAIGTTFGLNNIYLLKV-GKPGNKPAIFMDCGFHAREWISHAFCCQWFVREAVITY 185
Qy 174 -----GH-----NMWRKRSFYANNHCIGTDLNRNFASK 203
Db 186 GYESHMTFLNKLDFYVLPVNLIDGIYITWTKNRMWRKRSNAGTTCIGTDPNRF-DA 244
```



```

Best Local Similarity 36.0%; Pred. No. 4.6e-60;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

Qy 1 FQSQVLAALPRTSRQVQLNLTYYEIVLWQPVTDALIVKKKQVHFFVNASVDNVKA 60
Db 20 FEGEKVFVNVEDENHINIIRELASTTQIDFWKPDSTVQIKPHSTVDFRKAEDTVTVEN 79
Qy 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQVHSLNEIYSWIEPTTERHPD 120
Db 80 VLKQNELOQVKLISNLRNVVEAQDSRV---RATGHSYKYNKWTETIETATQVATENPA 136
Qy 121 MLTKIHIGSSPEKYPVLVLYKSGKEQAKNAIWDICGHHAREWISPAFCWFI----- 173
Db 137 LISRSVIGTTFEGRAIYLLKV-GRAGQNKPAIFMDCGFHAREWISPAFCWVREAVRTY 195
Qy 174 -----GH-----NRMWRKNSFYANNHCIGTDLNRPASK 203
Db 196 GREIQVTELDKLDFFVLPVLNIDGYIYTWTKSRFWKTRSTHTGSSCIGTDPNRF-DA 254
Qy 204 HWCEGASSSSCSTYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHIYFPYSY 263
Db 255 GWCEIGASRNPDCDTCYCPAAESEKETKALADFRNKLSSIKAYLTTHSYSSOMMIYFYSY 314
Qy 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDMIYDLGIKYSF 323
Db 315 AYKLGENNAELNALAKATVKEL-ASLHGTYKTYGPGATTIYPAGGGDDWAYDOGIYRYSF 373
Qy 324 T 324
Db 374 T 374

RESULT 10
US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARGHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011,769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-011-769A-56

Query Match 35.5%; Score 637.5; DB 4; Length 424;
Best Local Similarity 36.0%; Pred. No. 4.7e-60;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

Qy 1 FQSQVLAALPRTSRQVQLNLTYYEIVLWQPVTDALIVKKKQVHFFVNASVDNVKA 60
Db 29 FEGEKVFVNVEDENHINIIRELASTTQIDFWKPDSTVQIKPHSTVDFRKAEDTVTVEN 88
Qy 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQVHSLNEIYSWIEPTTERHPD 120
Db 89 VLKQNELOQVKLISNLRNVVEAQDSRV---RATGHSYKYNKWTETIETATQVATENPA 145
Qy 121 MLTKIHIGSSPEKYPVLVLYKSGKEQAKNAIWDICGHHAREWISPAFCWFI----- 173
Db 146 LISRSVIGTTFEGRAIYLLKV-GRAGQNKPAIFMDCGFHAREWISPAFCWVREAVRTY 204
Qy 174 -----GH-----NRMWRKNSFYANNHCIGTDLNRPASK 203
Db 205 GREIQVTELDKLDFFVLPVLNIDGYIYTWTKSRFWKTRSTHTGSSCIGTDPNRF-DA 263
Qy 204 HWCEGASSSSCSTYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHIYFPYSY 263
Db 264 GWCEIGASRNPDCDTCYCPAAESEKETKALADFRNKLSSIKAYLTTHSYSSOMMIYFYSY 323
Qy 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDMIYDLGIKYSF 323
Db 324 AYKLGENNAELNALAKATVKEL-ASLHGTYKTYGPGATTIYPAGGGDDWAYDOGIYRYSF 382
Qy 324 T 324
Db 383 T 383

RESULT 11
US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Baton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896

```

TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-649-5918-7

Query Match 35.2%; Score 632; DB 1; Length 417;
Best Local Similarity 38.2%; Pred. No. 1.8e-59;
Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;

Qy 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTDALIVKKQVHFFVNASVDVNVKA 60
Db 21 FDREKVPFVKLQNEKHASVLKNTQSIELDFWYDPAIDHDAVNMVDFRVSKEKESQTIQS 80

Qy 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEQYHSLNIEYSWIEFITERHP 119
Db 81 TLEQHKHYELIHDLEIEKQFDVDEIAGRHS---YAKNDWDKIVSWTEKMLEKHP 137

Qy 120 DMLTKIHIGSFKEYPLVYLKVSKEQTAKNAIWDCGIIHAREWISPAFCLWFI----- 173
Db 138 EMVSRIKIGSTVEDNPLYVLKI-GKDGKRAIFMDCGIIHAREWISPAFCQWFVYQATKS 196

Qy 174 -GHN-----RMWRKNRSFYANNHCITGDLNRNPFAS 202
Db 197 YGKNKIMTKLLDRMNFVLPVFNVDGYIWSWTQDRMRKNRSNQNSTCIGTDLNRNF-D 255

Qy 203 KHWCEGASSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHIVFPYS 262
Db 256 VSWDSSNTKPCNLVYRGPAPESEKETKAVTNFIRSHLSIKAYITFHYSQMLLIPYG 315

Qy 263 YTRSKSDHELSLVAESAARAIEKTSKNTRYTHGSGSETLYLAPGGGDDWIYDLGIKYS 322
Db 316 YTFKLPPNHQDLKVARIAIDAL-STRYETRYIYGPIASTIYKTSGSSLDWVYDLGIKHT 374

Qy 323 F 323
Db 375 F 375

RESULT 12
US-08-277-540-7
Sequence 7, Application US/08277540
Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Haasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689DIC1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-7

Query Match 35.2%; Score 632; DB 1; Length 417;
Best Local Similarity 38.2%; Pred. No. 1.8e-59;
Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;

Qy 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTDALIVKKQVHFFVNASVDVNVKA 60
Db 21 FDREKVPFVKLQNEKHASVLKNTQSIELDFWYDPAIDHDAVNMVDFRVSKEKESQTIQS 80

Qy 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEQYHSLNIEYSWIEFITERHP 119
Db 81 TLEQHKHYELIHDLEIEKQFDVDEIAGRHS---YAKNDWDKIVSWTEKMLEKHP 137

Qy 120 DMLTKIHIGSFKEYPLVYLKVSKEQTAKNAIWDCGIIHAREWISPAFCLWFI----- 173
Db 138 EMVSRIKIGSTVEDNPLYVLKI-GKDGKRAIFMDCGIIHAREWISPAFCQWFVYQATKS 196

Qy 174 -GHN-----RMWRKNRSFYANNHCITGDLNRNPFAS 202
Db 197 YGKNKIMTKLLDRMNFVLPVFNVDGYIWSWTQDRMRKNRSNQNSTCIGTDLNRNF-D 255

Qy 203 KHWCEGASSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHIVFPYS 262
Db 256 VSWDSSNTKPCNLVYRGPAPESEKETKAVTNFIRSHLSIKAYITFHYSQMLLIPYG 315

Qy 263 YTRSKSDHELSLVAESAARAIEKTSKNTRYTHGSGSETLYLAPGGGDDWIYDLGIKYS 322
Db 316 YTFKLPPNHQDLKVARIAIDAL-STRYETRYIYGPIASTIYKTSGSSLDWVYDLGIKHT 374

Qy 323 F 323
Db 375 F 375

RESULT 13
US-08-430-787A-7
Sequence 7, Application US/08430787A
Patent No. 5593674
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995

CLASSIFICATION: 514
PRIOR APPLICATION DATA: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Haesk, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-7

Query Match 35.2%; Score 632; DB 1; Length 417;
Best Local Similarity 38.2%; Pred. No. 1.8e-59;
Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;
Qy 1 FQSGVLAALPRTSRQVQLQNLTTTVEIVLWQPVTDADLVKKQVHFFVNASDNDVKA 60
Db 21 FDRKVFVRVQLQNEKHSVLKNTLSIELDFWYDAIHDIAVNTVDFRVSKESESQTIQS 80
Qy 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEQYHSLNBIYSWIEFITERHP 119
Db 81 TLEQKHIEYLIDHQEEIEKQDFVDKEIAGRHS---YAKYNDWDKIVSWTERKMLEKHP 137
Qy 120 DMLTKIHIGSSFYKPLYVLKVSKEQTAKNAIWDCGIIHAREWISPAFCLEWFI----- 173
Db 138 EMVSRKIGSTVEDNPLYVLKI-GKDGGERKAIFMDCGIIHAREWISPAFCQWVYQATKS 196
Qy 174 -GHN-----RMRKNSFYANNHCIGTDLNRNFS 202
Db 197 YGRNKIMTKLLDRNFFVLPVFNVDGYIWSQTRMRKNSRQNSCTCIGTDLNRNF-D 255
Qy 203 KHWCEGASSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHIYFPYS 262
Db 256 VSWDSSPTNKPCLVYRGPAPESEKETKAVTNFIRSHLSIKAYITFHSYSQMLLPYG 315
Qy 263 YTRSKSDHBELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGGDDWIYDLGIKYS 322
Db 316 YTFKLPPNHODLLKVARIAIDAL-STRYETRIYIGPIASTIYKTSGSSLDWVYDLGIKHT 374
Qy 323 F 323
Db 375 F 375

RESULT 14
US-07-649-591B-6
Sequence 6, Application US/07649591B
Patent No. 5206161
GENERAL INFORMATION:
APPLICANT: Dennis Drayna and Daniel Eaton
TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California

COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/649,591B
FILING DATE: 19910201
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Haesk, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/266-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-649-591B-6

Query Match 34.6%; Score 622; DB 1; Length 417;
Best Local Similarity 35.5%; Pred. No. 2.2e-58;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;
Qy 1 FQSGVLAALPRTSRQVQLQNLTTTVEIVLWQPVTDADLVKKQVHFFVNASDNDVKA 60
Db 21 FDRKVFVRVQLQNEKHSVLKNTLSIELDFWYDAIHDIAVNTVDFRVSKESESQTIQS 80
Qy 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEQYHSLNBIYSWIEFITERHP 119
Db 81 ALQONKHIEYLIDHQEEIEKQDFVDKEIAGRHS---YAKYNNWEKIVAWTEKMDKYP 137
Qy 120 DMLTKIHIGSSFYKPLYVLKVSKEQTAKNAIWDCGIIHAREWISPAFCLEWFI----- 173
Db 138 EMVSRKIGSTVEDNPLYVLKI-GERNERKAIFMDCGIIHAREWISPAFCQWVYQATKT 196
Qy 174 -GHN-----RMRKNSFYANNHCIGTDLNRNFS 202
Db 197 YGRNKIMTKLLDRNFFVLPVFNVDGYIWSQTRMRKNSRQNSCTCIGTDLNRNF-N 255
Qy 203 KHWCEGASSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHIYFPYS 262
Db 256 ASWNSPTNTPCADNTRGAPASEKETKAVTNFIRSHLSIKAYITFHSYSQMLLPYG 315
Qy 263 YTRSKSDHBELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGGDDWIYDLGIKYS 322
Db 316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRIYIGPIESTIYPISSGSLDWDYDLGIKHT 374
Qy 323 F 323
Db 375 F 375

RESULT 15
US-08-277-540-6
Sequence 6, Application US/08277540
Patent No. 5474901
GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd

CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/952-9896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 417 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-277-540-6

Query Match 34.6%; Score 622; DB 1; Length 417;
Best Local Similarity 35.5%; Pred. No. 2.2e-58;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;
Qy 1 PQSQVLAALPRTSRQVQLNLTITTEIVLVQPTADLIIVKKQVHFFVNASDVNDVKA 60
Db 21 FDRKVPVRVKODEKQADIIKDLAKTNELDFWPCATHHVAANMMVDFRVSEKESQAIQS 80
Qy 61 HLNVSIGIPCSVLLADVEDLIQQOIS-NDTVSPRASASYEOYHSLNEIYSWIEFITERHP 119
Db 81 ALDQKMHYELIHDLQEEIEKQFDVKEDI PGRHS---YAKTNWEKIVAWTEKMDKIP 137
Qy 120 DMLTKIHIGSFERYPLVYLVKSGKEQTAKNAIWDGCIHAREWISPAFCLWFI----- 173
Db 138 EMVSRKIGSTVEDNPLVVLKI-GEKNERRKAI FMDCGIHAREWVSPAFQCFVYQATKT 196
Qy 174 -----GH-----NRMWRKNSFYANNHCITDLDNRNPAS 202
Db 197 YGRNKIMTKLDRMNFYILPVFVNDGYIWSWTKRNMWRKNSKQNSKICIGTDLNRNF-N 255
Qy 203 KHWCEGASSSCSEFTYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSOHIWVPS 262
Db 256 ASWNSIPNTNDFCADNHYGSAPESEKETKATVNFIRSHLNEIKVYITPHSYSQMLLPYG 315
Qy 263 YTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHSSETLYLAPGGDDWIYDLGIKYS 322
Db 316 YTSKLPPNHEDLAKVAKIGTDVL-STRVETRYIYGIPISTIYPIGSSLDWAYDLGIKHT 374
Qy 323 F 323
Db 375 F 375

Search completed: January 1, 2004, 19:33:10
Job time : 21.8163 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:32:17 ; Search time 48.3361 Seconds
(without alignments)
1403.024 Million cell updates/sec

Title: US-09-980-881a-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1699.5	94.5	423	12	US-10-379-836-17
2	1615.5	89.8	423	12	US-10-379-836-2
3	1487.5	82.7	422	12	US-10-379-836-18
4	1466.5	81.6	422	12	US-10-379-836-16
5	816.5	45.4	211	9	US-09-925-302-467
6	637.5	35.5	402	12	US-10-379-836-20
7	622	34.6	417	12	US-10-341-434-188
8	622	34.6	417	15	US-10-229-546-2
9	622	34.6	417	15	US-10-229-546-9
10	619.5	34.5	716	10	US-09-910-059-125
11	612	34.0	437	12	US-10-274-639-12
12	612	34.0	437	14	US-10-200-344-10
13	569	31.6	374	9	US-09-888-615-61
14	569	31.6	444	16	US-10-176-306-74
15	554	30.8	399	15	US-10-200-910-8

16	549	30.5	613	10	US-09-910-059-113	Sequence 113, App
17	534.5	29.7	350	14	US-10-200-344-12	Sequence 12, Appl
18	527.5	29.3	436	15	US-10-200-910-6	Sequence 6, Appl1
19	519.5	28.9	436	14	US-10-076-535-2	Sequence 2, Appl1
20	511.5	28.4	417	9	US-09-923-779-150	Sequence 150, App
21	507.5	28.2	428	9	US-09-925-297-528	Sequence 528, App
22	501.5	27.9	298	12	US-10-379-836-19	Sequence 19, Appl
23	493.5	27.4	406	9	US-09-925-297-517	Sequence 517, App
24	493.5	27.4	419	11	US-10-345-680-65	Sequence 65, Appl
25	491	27.3	421	11	US-09-946-374-25	Sequence 234, App
26	491	27.3	421	12	US-10-006-130A-234	Sequence 234, App
27	491	27.3	421	12	US-10-015-387A-234	Sequence 234, App
28	491	27.3	421	12	US-10-199-672-308	Sequence 308, App
29	491	27.3	421	12	US-10-006-172A-234	Sequence 234, App
30	491	27.3	421	12	US-10-187-749-308	Sequence 308, App
31	491	27.3	421	12	US-10-184-457-308	Sequence 308, App
32	491	27.3	421	12	US-10-184-643-308	Sequence 308, App
33	491	27.3	421	12	US-10-196-747-308	Sequence 308, App
34	491	27.3	421	12	US-10-015-392A-234	Sequence 234, App
35	491	27.3	421	12	US-10-017-253A-234	Sequence 234, App
36	491	27.3	421	12	US-10-173-689-308	Sequence 308, App
37	491	27.3	421	12	US-10-173-690-308	Sequence 308, App
38	491	27.3	421	12	US-10-173-691-308	Sequence 308, App
39	491	27.3	421	12	US-10-173-692-308	Sequence 308, App
40	491	27.3	421	12	US-10-173-694-308	Sequence 308, App
41	491	27.3	421	12	US-10-173-698-308	Sequence 308, App
42	491	27.3	421	12	US-10-173-707-308	Sequence 308, App
43	491	27.3	421	12	US-10-173-707-308	Sequence 308, App
44	491	27.3	421	12	US-10-174-569-308	Sequence 308, App
45	491	27.3	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1

US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Query Match	94.5%	Score 1699.5	DB 12	Length 423
Best Local Similarity	89.8%	Pred. No. 7.7e-166		
Matches 324	Conservative 0	Mismatches 0	Indels 37	Gaps 1
Qy	1	FQSGQVLAALPRTSRQVQLQNLTTTYYEIVLWQPVTDALIVKKKQVHFFVNASDVNDVKA	60	
Db	23	FQSGQVLAALPRTSRQVQLQNLTTTYYEIVLWQPVTDALIVKKKQVHFFVNASDVNDVKA	82	
Qy	61	HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASASYEQYHSLNEYISYIEFTTERHPD	120	
Db	83	HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASASYEQYHSLNEYISYIEFTTERHPD	142	
Qy	121	MLTKIHGSSFEKYPVLYLVKVSKEQTAKNAIWDGCIHAREWISPAFLWFIGH-----	175	
Db	143	MLTKIHGSSFEKYPVLYLVKVSKEQTAKNAIWDGCIHAREWISPAFLWFIGHTQFY	202	
Qy	176	-----NRMRKNRSFYANNHCIGTDLNRNFASK	203	


```
Best Local Similarity 76.7%; Pred. No. 7e-142;
Matches 277; Conservative 22; Mismatches 25; Indels 37; Gaps 1;

QY 1 FQSGQVLAALPRTSRQVQLQLNTTYYEIVLWQPVTDADLVKKQVHFFVNASDQVNVKA 60
Db 22 FQSGHVLALPRTSRQVQLQLNTTYYEIVLWQPVTDADLVKKQVHFFVNASDQVNVKA 81
QY 61 HLNVSIPCSVLADVEDLIQQOISNDTVSPRASASYEYQYHSLNIEYSWIEFITERHPD 120
Db 82 YLNASRIPFVNLNNVEDLIQQOISNDTVSPRASASYEYQYHSLNIEYSWIEFITEQHPD 141
QY 121 MLTKIHGSSFEKYPLYVLKVSKEQTAKNAIWDGCIHAREWISPAFCLEWFGH----- 175
Db 142 MLQKIYGSSYEKYPLYVLKVSKEHVRKNAIWDGCIHAREWISPAFCLEWFGHYTQFH 201
QY 176 -----NRMKRNRSFYANNHCIGTDLNRNFASK 203
Db 202 GKENTYTRLLRHVDYFIMPVNVGDYDTWKKNRMKRNRSVHMNRNRCVGTDLNRNFASK 261
QY 204 HWCEGASSSCSETCYGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHIIVFPYSY 263
Db 262 HWCEKASSFCSETCYGLYPESEPEVKAVADFLRRNINNIHAKAYISMHSYSQOILFPYSY 321
QY 264 TRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSF 323
Db 322 NRSKSDHELSLVASEAVRAIESINKNTRYTHGSGSESLYAPGGDDWIYDLGIKYSF 381
QY 324 T 324
Db 382 T 382

RESULT 5
US-09-925-302-467
; Sequence 467, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 467
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-467

Query Match 45.4%; Score 816.5; DB 9; Length 211;
Best Local Similarity 90.6%; Pred. No. 1.7e-75;
Matches 154; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

QY 163 WISPAFLWFI-----GHNRMKRNRSFYANNHCIGTDLNRNFASKHWCEGASSSS 214
Db 3 WIS-MLCRLWMMVMNYSWKKNRMKRNRSFYANNHCIGTDLNRNFASKHWCEGASSSS 61
QY 215 CSSTYCYGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHIIVFPYSYTRSKSDHEEL 274
Db 62 CSSTYCYGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHIIVFPYSYTRSKSDHEEL 121
QY 275 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 324
Db 122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 171

RESULT 6
US-10-379-836-20
```

```
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAPI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-20

Query Match 35.5%; Score 637.5; DB 12; Length 402;
Best Local Similarity 36.0%; Pred. No. 1.1e-56;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

QY 1 FQSGQVLAALPRTSRQVQLQLNTTYYEIVLWQPVTDADLVKKQVHFFVNASDQVNVKA 60
Db 7 FEGEKVRVNVDENNHIIRELASTQIDFWKPDSTQIKPHSTVDFRVKAEDTIVTVEN 66
QY 61 HLNVSIPCSVLADVEDLIQQOISNDTVSPRASASYEYQYHSLNIEYSWIEFITERHPD 120
Db 67 VLKQNELQYKVLISLRNVVEAQFDSRV---RATGSEYKYNKWEITEATQVATENPA 123
QY 121 MLTKIHGSSFEKYPLYVLKVSKEQTAKNAIWDGCIHAREWISPAFCLEWFI----- 173
Db 124 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQMFVREAVRTY 182
QY 174 -----GH-----NRMKRNRSFYANNHCIGTDLNRNFASK 203
Db 183 GREIQVTELLNKLDYFVLPVNLIDGYITWTKSRFRKTRSTHTGSCICTDPRNPF-DA 241
QY 204 HWCEGASSSCSETCYGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHIIVFPYSY 263
Db 242 GWCEIGASRNPCEITYCGPAASEKETKALADFIRNKLSSIKAYLTIHISYSQMMIYPYSY 301
QY 264 TRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSF 323
Db 302 AYKLGNNNAELNALAKATVKEL-ASLHGTYTYGPGATTIYPAGGGDDWAYDQGIKYSF 360
QY 324 T 324
Db 361 T 361

RESULT 7
US-10-341-434-188
; Sequence 188, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 188
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-188
```

```
Query Match 34.6%; Score 622; DB 12; Length 417;
Best Local Similarity 35.5%; Pred. No. 4.6e-55;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;

Qy 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQVPTADLVKKQVHFFVNASDNDVKA 60
Db 21 FDEKVRVFPQDEKQADIKDLAKTNELDFWYPGATHVAAANMVDVFRVSEKESQAIQS 80

Qy 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEQYHSLNEIYSWIFETERHP 119
Db 81 ALDONKMHYEILLHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKQMDKYP 137

Qy 120 DMLTKIHGSSFEKYPFLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCIWF- 173
Db 138 EMVSRIKIGSTVEDNPPLYVLI-GEKNERRKAIKFMDCGIIHAREWVSPAFQWVYQATKT 196

Qy 174 -----GH-----NRMWRKNRSFYANNHCIGTDLNRNFAS 202
Db 197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTQNRWKRKNQSKQNSKCIOTDLNRNF-N 255

Qy 203 KHCEBEGASSSSCSETVCGLYPSEPEVKAVASFLRNINQIKAYISMHSYSQHIYPPYS 262
Db 256 ASWNSIPNTNDPCADNVRGSAPESEKETKAVTNFIRSHLNEIKVYITFFHSYSQWMLFPYG 315

Qy 263 YTRSKDKHEELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDMIYDLGIKYS 322
Db 316 YTSKLPNHDHAKVAKIGTDVL-STRYETRYIYGPSTIYPIYSGSSLDWAYDLGIKHT 374

Qy 323 F 323
Db 375 F 375

RESULT 8
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299. A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MP101-156PIRNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-546-2

Query Match 34.6%; Score 622; DB 15; Length 417;
Best Local Similarity 35.5%; Pred. No. 4.6e-55;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;

Qy 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQVPTADLVKKQVHFFVNASDNDVKA 60
Db 21 FDEKVRVFPQDEKQADIKDLAKTNELDFWYPGATHVAAANMVDVFRVSEKESQAIQS 80

Qy 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEQYHSLNEIYSWIFETERHP 119
Db 81 ALDONKMHYEILLHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKQMDKYP 137

Qy 120 DMLTKIHGSSFEKYPFLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCIWF- 173
Db 138 EMVSRIKIGSTVEDNPPLYVLI-GEKNERRKAIKFMDCGIIHAREWVSPAFQWVYQATKT 196

Qy 174 -----GH-----NRMWRKNRSFYANNHCIGTDLNRNFAS 202
Db 197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTQNRWKRKNQSKQNSKCIOTDLNRNF-N 255

Qy 203 KHCEBEGASSSSCSETVCGLYPSEPEVKAVASFLRNINQIKAYISMHSYSQHIYPPYS 262
Db 256 ASWNSIPNTNDPCADNVRGSAPESEKETKAVTNFIRSHLNEIKVYITFFHSYSQWMLFPYG 315

Qy 263 YTRSKDKHEELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDMIYDLGIKYS 322
Db 316 YTSKLPNHDHAKVAKIGTDVL-STRYETRYIYGPSTIYPIYSGSSLDWAYDLGIKHT 374

Qy 323 F 323
Db 375 F 375

RESULT 9
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299. A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MP101-156PIRNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-546-9

Query Match 34.6%; Score 622; DB 15; Length 417;
Best Local Similarity 35.5%; Pred. No. 4.6e-55;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;

Qy 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQVPTADLVKKQVHFFVNASDNDVKA 60
Db 21 FDEKVRVFPQDEKQADIKDLAKTNELDFWYPGATHVAAANMVDVFRVSEKESQAIQS 80

Qy 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEQYHSLNEIYSWIFETERHP 119
Db 81 ALDONKMHYEILLHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKQMDKYP 137

Qy 120 DMLTKIHGSSFEKYPFLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCIWF- 173
Db 138 EMVSRIKIGSTVEDNPPLYVLI-GEKNERRKAIKFMDCGIIHAREWVSPAFQWVYQATKT 196

Qy 174 -----GH-----NRMWRKNRSFYANNHCIGTDLNRNFAS 202
Db 197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTQNRWKRKNQSKQNSKCIOTDLNRNF-N 255

Qy 203 KHCEBEGASSSSCSETVCGLYPSEPEVKAVASFLRNINQIKAYISMHSYSQHIYPPYS 262
Db 256 ASWNSIPNTNDPCADNVRGSAPESEKETKAVTNFIRSHLNEIKVYITFFHSYSQWMLFPYG 315

Qy 263 YTRSKDKHEELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDMIYDLGIKYS 322
Db 316 YTSKLPNHDHAKVAKIGTDVL-STRYETRYIYGPSTIYPIYSGSSLDWAYDLGIKHT 374

Qy 323 F 323
Db 375 F 375
```

RESULT 10
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; TITLE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 125
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: humanised pre-pro HCPB-linker-Pd sequence
US-09-910-059-125

Query Match 34.5%; Score 619.5; DB 10; Length 716;
Best Local Similarity 35.2%; Pred. No. 1.8e-54;
Matches 127; Conservative 69; Mismatches 122; Indels 43; Gaps 6;

QY 1 FQSGQVLAALPRTSRQVQLVQLTYYEIVLWQPVTDLVKVKQVHFFVNASDVNVKA 60
DB 22 FEGEGRVFNVDENHNIIRLASTQIDFWKPSVTQIKPHSTVDFRVKAEDTYTVEN 81
QY 61 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASASYEYQYHSLNEIYSWIEPIRHPD 120
DB 82 VLKQNELOYKVLISLNRVNVEAFDSRV---RATGSHYKYNKWEIETQVATENPA 138
QY 121 MLTKIHGSSFEKPYLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCMLWFI----- 173
DB 139 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 197
QY 174 -----GH-----NRMWRKRSFYANNHCITGDLNRNPA 203
DB 198 GREIQVTELDKLDYVLPVNLIDGYIYTWKSRFRWKRTRSTHTGSSCGTDFNRF-DA 256
QY 204 HWCCEGASSSSCSEYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIYFPYSY 263
DB 257 GWCEIGASRNPCDETCGPAESEKETKALADIRNKLISIKAYLTHISYQMMIYPIYSY 316
QY 264 TRSKDKHELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSF 323
DB 317 AVKLGENNAELNALAKATVKEL-ASLHGTYKTYGPGATTIYPSAGTSKWAYDQGIYRYSF 375
QY 324 T 324
DB 376 T 376

RESULT 11
US-10-274-639-12
; Sequence 12, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DELEGEANE, Angelo M.; GANDHI, Ameena R.
; APPLICANT: HAFALIA, April J.A.; LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.

APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
APPLICANT: NGUYEN, Damiel B.; LEE, Ernestine A.
APPLICANT: KHAN, Farrah A.; YUE, Henry
APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
APPLICANT: YANG, Junming; THANGAVELU, Kavitha
APPLICANT: DING, Li; KEARNEY, Liam
APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
APPLICANT: BURFORD, Neil; WALIA, Narinder K.
APPLICANT: LAL, Preeti G.; LEE, Sally
APPLICANT: TODD, Stephen; LO, Terence P.
APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
APPLICANT: AZIMZAI, Yalda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,717
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/225,988
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 60/227,568
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 12
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

Query Match 34.0%; Score 612; DB 12; Length 437;
Best Local Similarity 35.4%; Pred. No. 5.2e-54;
Matches 129; Conservative 62; Mismatches 127; Indels 46; Gaps 7;

QY 1 FQSGQVLAALPRTSRQVQLVQLTYYEIVLWQPVTDLVK--KKQVHFFVNASDVNV 58
DB 37 YAGDKVIRPIKTEEBAYALKKISYQLKVDLWQPSISYVSEGTVDVHIPQNGSRA--L 94
QY 59 KAHLNVSGIPCSVLLADVEDLIQQOISNDTVSPRASAS--YVEQYHSLNEIYSWIEFITE 116
DB 95 LAFLOEANIQQYKVLIEDLQKTEKSSLHTQRRNRSLSGYNYEVYHSLBEEIQNMHHLNK 154
QY 117 RHPDMLTKIHGSSFEKPYLYLVKVSKEQTAKNAIWDGCIHAREWISPAFCMLWFI--- 173
DB 155 THSGLIHMSIGRSYBGRSLFILK--GRSRLKRAVWIDGCIHAREWIGPAFCQWFVKEA 213
QY 174 -----GHNRMWRKRSFYANNHCITGDLNRN 199
DB 214 LLTYKSDPAMRKMLNHLFYFVFNVDGYHFSWTNDRFWKTRSNRSPRCRGVDNRN 273
QY 200 FASKHWCCEGASSSSCSEYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIYF 259
DB 274 WKYK-WCEGASNHPCDDTYCGFPFPESEPEVKAVANFLRKRKHIRAYLSFHAYAOMLLY 332
QY 260 PYSYTRSKDKHELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGI 319
DB 333 PYSYKYATTIPNRCVESAAYKAVNALQSV-YGVRYRYGPASTTLYVSSSSMDWAYNGI 391
QY 320 KYSF 323
DB 320 KYSF 323

```
Db 392 PYAF 395

RESULT 12
US-10-200-344-10
; Sequence 10, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020173641a1 Human Carboxypeptidases and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 437
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-200-344-10

Query Match 34.0%; Score 612; DB 14; Length 437;
Best Local Similarity 35.4%; Pred. No. 5.2e-54;
Matches 129; Conservative 62; Mismatches 127; Indels 46; Gaps 7;

Qy 1 FQSGVLAALPRTSRQVQVTLQNTTVEIVLWQVTDADLVK--KKQVHFFVNASVDNV 58
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 YAGDKVIRFPKTEEEAYALKKISYQLKVDLWQPSSISYVSEGTVDVHIPONGSRA--L 94
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 59 KAHNVSGIPCSVLLADVEDLIQOQISNDTVSPRAS--YEQYHSLNEIYSWIEITE 116
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 LAFLQENAIQKVLIEDLQKTLKGGSLHTQNRNRSLSGYNVEYVYHLSLEIQNMHHLNK 154
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 117 RHPDMLTKIHGSSFEKYPLYVLKVSKEQTAKNAIWDGCIHAREWISPAFCILWFI--- 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 THSLIHFSGRSGYEGRSFILKL--GERSLKEAVWDGCIHAREWIGPAFCWFVKEA 213
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 174 -----GHNRMWRKNRSFYANNHCIGTDLNRN 199
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 LLTYKSDPAMKMLNHLFYIMPVNDGYHFSWTNDRFWKTRSRNRSRRCRGVDANRN 273
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 200 FASKHWCCEGASSSCSCSETCYGLYPESEPEKAVASFLRNINQIKAYISMHSYSQHIVF 259
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 WKVK-WCDEGASMHPCDDTYGPPFSEPEKAVANFLRKHRIAYLSFHAYAOQLLY 332
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 260 PYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHSETLYLAPGGGDDWIYDLGI 319
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 PYSKYATIPNRCVESAAYKANVALQSV-YGVRYRGPASTTLYVSSGSDMDWAYKNGI 391
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 320 KYSP 323
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 PYAF 395

RESULT 13
US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA

; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-61

Query Match 31.6%; Score 569; DB 9; Length 374;
Best Local Similarity 39.0%; Pred. No. 1.1e-49;
Matches 110; Conservative 51; Mismatches 81; Indels 40; Gaps 5;

Qy 81 QQOISNDTVSPRASV-YEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPLYVL 139
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29 RQIVDKSVSPWSLETYSYNIYHPMGEIYEWMEISEKYEKVVTOHFLGVTYTHPMYLL 88
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 140 KVSKEQTAKNAIWDGCIHAREWISPAFCILWFI----- 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 KISQPSGNPKKIIMDCGIHAREWIAPAFCWFVKEILQNHKDNSSIRKLLRLNLDYVLP 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 174 -----GH-----NRMWRKNRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETCYGL 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 VLNIDGYIYTWTDRLWRKSRSPHNNGTCFGTDLNRNF--NASWCSIGASRNCQDTFCGT 207
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 223 YPSEPEKAVASFLRNINQIKAYISMHSYSQHIVFPYSTRSKSDHELSLVASEAV 282
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 GPVSEPEKAVASFLRNINQIKAYISMHSYSQHIVFPYSTRSKSDHELSLVASEAV 267
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 283 RAIEKTSKNTRYTHGHSETLYLAPGGGDDWIYDLGIKYSFT 324
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 NAL-KAKYGTNYRVGSADILYASSGSSRDWARDIGIPFSYT 308
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-10-176-306-74
; Sequence 74, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Curtiss, Kory A. J.
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
; FILE REFERENCE: 10448-195001
; CURRENT APPLICATION NUMBER: US/10/176,306
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 10/001,137
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: PCT/US01/45291
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/248,362
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,331
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,365
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/250,077
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,327
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,176
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 10/023,617
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/49416
; PRIOR FILING DATE: 2001-12-18
```

```
; PRIOR APPLICATION NUMBER: 60/256,249
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,405
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 10/083,248
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/US01/46717
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242,324
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,518
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-176-306-74

Query Match      31.6%; Score 569; DB 16; Length 444;
Best Local Similarity 39.0%; Pred. No. 1.4e-49;
Matches 110; Conservative 51; Mismatches 81; Indels 40; Gaps 5;

Qy 81 QOQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEKYPYLV 139
Db 159 RQEVDSKVSFWSLETYSYNIYHPMGEIYEMWREISEKYKEVVTQHPGLGVTYETHPIYL 218
Qy 140 KVSQKEQTAKNAIWIDCGIHAREWISPAFCLWFI----- 173
Db 219 KISQPSGNPKKIIMWDCGIIHAREWIAFPCQWFVKEILQNHKDSRIKLLRNLDFFVLP 278
Qy 174 -----GH-----NRMRKNRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSCSYTCGL 222
Db 279 VLNIDGVIYTWTDRLWRKRSRPHNGTCFTGDLNRNF-NASWCSTGASRNCQDQTFCGT 337
Qy 223 YPSEPEVKAVASFLRRNINQIKAYISMHSYSQHVFPYSTRSKSDHEELSLVASEAV 282
Db 338 GPVSEPETKAVASFIESKDDILCFLTHMSYQGLILTPYGYTKNKSNNHPMIQVGQKAA 397
Qy 283 RAJEKTSKNTRYTHGHSETLYLAPGGDDWIYDLGIKYSFT 324
Db 398 NAL-KAKYGTNYRVGSSADILYASSGSSRDRWARDIGIPPSYT 438

RESULT 15
US-10-200-910-8
; Sequence 8, Application US/10200910
; Publication No. US20030009021A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030009021A1el Human Proteases and
; FILE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0086-USA
; CURRENT APPLICATION NUMBER: US/10/200,910
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US/09/710,099
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: US 60/165,260
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 399
; TYPE: PRT

; ORGANISM: homo sapiens
; US-10-200-910-8

Query Match      30.8%; Score 554; DB 15; Length 399;
Best Local Similarity 39.0%; Pred. No. 4.2e-48;
Matches 130; Conservative 60; Mismatches 123; Indels 20; Gaps 10;

Qy 1 FOSGQVLAALPRTSRQVQLQNL---TTTYEIVLWQ-PVTADLVKKKQVHFFNADVDN 57
Db 37 FTGDQVLRVLAKDEKQLSLGLDLEGLKPKQKVPFMRGPAPSLPVDNR-----VPFSELKD 91
Qy 58 VKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEF 113
Db 92 IKAYLESHGLAYSIMIKDIQVLDDEBERQAWAKSRRLERSTNSFSYSYHTLEIYSWIDN 151
Qy 114 ITERHPDMLTKIHIGSSFKEKYPYLVKVS--GKEQTAKNAIWIDCGIHAREWISPAFCLW 171
Db 152 FVMEHSDIVSKIQIGNSFENQSLVLKFTSGGSRHP---AIWIDTGIHSREWITHTATGIW 208
Qy 172 FIGHNRMRKNRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSCSYTCGLYPESEPEVK 231
Db 209 TANKNRLWRKNKSIIRPGIFCIGVDLNRNWK-S-GFGGNGSNSNPCSETYHGPSPQSEPEVA 267
Qy 232 AVASFLRRNINQIKAYISMHSYSQHVFPYSTRSKSDHEELSLVASEAVRAIEKTSKN 291
Db 268 AIVNFITAHGN-FKALISHSYSQMLMYPYGRLLLEPVSNQRELYDLAKDAVEALYKV-HG 325
Qy 292 TRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 324
Db 326 IEVIFGSISTLLVASGITVDWAYDSGIKYAFS 358

Search completed: January 1, 2004, 19:50:58
Job time : 50.3361 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:25:56 ; Search time 23.286 Seconds
(without alignments)
1395.903 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQL.....IKYFTSNPPVKLLPLSLK 338
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1699.5	94.5	423	2 A41204	carboxypeptidase B
2	691.5	38.5	415	2 A32129	carboxypeptidase B
3	632	35.2	417	1 A34487	carboxypeptidase A
4	622	34.6	417	2 A43929	carboxypeptidase A
5	618	34.4	416	1 A42332	carboxypeptidase B
6	586	32.6	306	1 CPBOB	carboxypeptidase B
7	552	30.7	309	2 A38395	mast cell carboxyp
8	549.5	30.6	417	2 A32128	carboxypeptidase A
9	536.5	29.8	419	1 CPBOA	carboxypeptidase A
10	532.5	29.6	419	1 CPRTA	carboxypeptidase A
11	511.5	28.4	417	2 A56171	carboxypeptidase A
12	493.5	27.4	419	1 S29127	carboxypeptidase A
13	443	24.6	528	2 T33527	hypothetical prote
14	403	22.4	323	2 T20507	hypothetical prote
15	401	22.3	667	2 T33526	hypothetical prote
16	400.5	22.3	527	2 T27572	hypothetical prote
17	375.5	20.9	303	1 CPCVB	carboxypeptidase B
18	372.5	20.7	455	2 T33572	hypothetical prote
19	364	20.2	666	2 T24170	hypothetical prote
20	300	16.7	430	2 S48976	ECM14 protein - ye
21	286.5	15.9	424	2 S17571	carboxypeptidase T
22	283	15.7	451	2 T20723	carboxypeptidase (
23	252	14.0	497	2 T40260	carboxypeptidase p
24	250.5	13.9	373	2 T26030	hypothetical prote
25	184	10.2	999	2 T36021	probable zinc-bind
26	142.5	7.9	1446	2 T30916	carboxypeptidase D
27	140.5	7.8	558	2 S75104	hypothetical prote
28	116.5	6.5	707	2 T40070	origin recognition
29	109.5	6.1	985	2 T29910	hypothetical prote

30	108.5	6.0	985	2 D82776	pyruvate dehydroge
31	106.5	5.9	1389	2 I50090	carboxypeptidase 9
32	105	5.8	491	2 B96739	hypothetical prote
33	101	5.6	663	2 T37772	telomere length re
34	101	5.6	1216	2 H84629	hypothetical prote
35	99	5.5	376	1 E69957	gamma-D-glutamyl-L
36	99	5.5	444	2 S51199	tryptophan 5-monoo
37	97.5	5.4	355	2 C83850	gamma-D-glutamyl-L
38	97	5.4	444	1 S10489	tryptophan 5-monoo
39	94	5.2	477	2 S16383	carboxypeptidase E
40	94	5.2	1160	2 I40589	paraspotal crystal
41	94	5.2	1587	2 AB2012	hypothetical prote
42	93.5	5.2	318	2 S52424	homeodomain protei
43	93	5.2	1451	2 A36468	SP16 protein - yea
44	92.5	5.1	986	2 E90596	restriction-modifi
45	92.5	5.1	986	2 H90565	restriction modifi

ALIGNMENTS

RESULT 1

A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N;Alternate names: plasma carboxypeptidase B
C;Species: Homo sapiens (man)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C;Accession: A41204
R;Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A;Title: Isolation, molecular cloning, and partial characterization of a novel carboxypep
A;Reference number: A41204; MUID:92042093; PMID:1939207
A;Accession: A41204
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <EAT>
A;Cross-references: GB:M75106; NID:gl89686; PID:gl89687
C;Genetics:
A;Gene: GDB:CPB2
A;Cross-references: GDB:129546; OMIM:212070
A;Map position: 13q14.11-13q14.11
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-114/Domain: activation peptide #status predicted <ACP>
F;115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Query Match 94.5%; Score 1699.5; DB 2; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.1e-133;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

Qy	1	FQSGQVLAALPRTSRQVQLNLT	TTTYEIVLWQPTADLVKKQVHFFVNASDV	DNVKA	60
Db	23	FQSGQVLAALPRTSRQVQLNLT	TTTYEIVLWQPTADLVKKQVHFFVNASDV	DNVKA	82
Qy	61	HLNVGIPCSVLADVEDLIQQQI	SNDTSPRASASYEQYHSLNIYSWIEFITE	RHPD	120
Db	83	HLNVGIPCSVLADVEDLIQQQI	SNDTSPRASASYEQYHSLNIYSWIEFITE	RHPD	142
Qy	121	MLTKIHIGSFEKYPYLVKVSKEQTAKNAI	WIDCGIHAREWISPAFLWFIGH	-----	175
Db	143	MLTKIHIGSFEKYPYLVKVSKEQTAKNAI	WIDCGIHAREWISPAFLWFIGH	TFQFY	202
Qy	176	-----	NRMRKNSFYANNHCIGTDLNRNFASK	203	
Db	203	GIITQYTNLLRLVDFVMPVNVVDGYDYS	WKQMRKNSFYANNHCIGTDLNRNFASK	262	
Qy	204	HWCEGASSSSCSSETYCGLYPESEPEVKA	VASFLLRRNIQIKAYISMHSYSQHVFPYSY	263	
Db	263	HWCEGASSSSCSSETYCGLYPESEPEVKA	VASFLLRRNIQIKAYISMHSYSQHVFPYSY	322	
Qy	264	TRSKSKDHELSLVASEAVRAIEKTSNTRY	THIGHSETLYLAPGGDDWIDYLGIKYSF	323	

Db 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 382
Qy 324 T 324
Db 383 T 383

RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C/Accession: A32129; S17543
R/Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A/Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparison of the two genes.
A/Reference number: A92693; MUID:89034324; PMID:3182872
A/Accession: A32129
A/Molecule type: DNA
A/Residues: 1-415 <CLA>
A/Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R/Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A/Title: Distribution of manganese in rat pancreas and identification of its primary binding site.
A/Reference number: S17543; MUID:91378950; PMID:1898371
A/Accession: S17543
A/Molecule type: protein
A/Residues: 109-116, 'X', 118-130 <XOD>
C/Genetics: CPB
A/Gene: CPB
A/Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C/Superfamily: carboxypeptidase
C/Keywords: hydrolase; metallo-carboxypeptidase
F;354,376/Active site: Tyr, Glu #status predicted

Query Match 38.5%; Score 691.5; DB 2; Length 415;
Best Local Similarity 40.2%; Pred. No. 9.1e-50;
Matches 145; Conservative 59; Mismatches 114; Indels 43; Gaps 6;

Qy 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTAADLVKKQVHFFVNASDNDVKA 60
Db 20 FDGNRVYRVSHGDDHVNLIQELANTKEIDFPWKPDSTQVQKPLTTVDHFVKAEDVADVEN 79

Qy 61 HLNVSIGPCSVLLADVEDLIQQOIS-NDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 120
Db 80 FLEENEVHYEVLISNVRALESQDSHT--RASGHSYTYKYNKWTIEAWIQQVATDNDP 136

Qy 121 MLTKIHGSSPEKYPYLVKVSQEQTAQNAIWDGCIHAREWISPAFCFLWFI----- 173
Db 137 LVTSQVGTTFEGRNMYLVKI-GKTRPNKPAIFIDCGFHAREWISPAFCQMFVREAVRTY 195

Qy 174 -----GH-----NMWRKNRSFYANNHCIGTDLNRNFSK 203
Db 196 NQEIHMKQLLDLDFYVLPVNNIDGVYVTTKDRMWRKTRSTWAGSSCLGVRPNRF-NA 254

Qy 204 HWCEGSSSSCSSTYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSY 263
Db 255 GWCVGASRSPCSSTYCGPAPSEKETKALADFIIRNLSTIKAYITHTSYSQMMLYPYSY 314

Qy 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 323
Db 315 DYKLPENYEELNALVKAANEL-ATLHGTYKTYGPGATTYPAGGSDSDSYDGIKYSF 373

Qy 324 T 324
Db 374 T 374

RESULT 3
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C/Accession: A34487
R/Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A/Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member of the carboxypeptidase A family.
A/Reference number: A34487; MUID:90062123; PMID:2584208
A/Accession: A34487
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-417 <REY>
A/Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:g309135
C/Superfamily: carboxypeptidase
C/Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

Query Match 35.2%; Score 632; DB 1; Length 417;
Best Local Similarity 38.2%; Pred. No. 8.2e-45;
Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;

Qy 1 FQSGQVLAALPRTSRQVQLNLTYYEIVLWQPVTAADLVKKQVHFFVNASDNDVKA 60
Db 21 FDREKVRVQLQNEKHAASVLKLTQSIEDFWYPDHAIIVNMTVDFRVSEKESQTIQS 80

Qy 61 HLNVSIGPCSVLLADVEDLIQQOIS-NDTVSPRASASYEQYHSLNEIYSWIEFITERHP 119
Db 81 TLSEQKHIVHEILLHDLQEEIEKQFDVKDEIAGRHS--YAKYNDWKIVSWTEKMLEKHP 137

Qy 120 DMLTKIHGSSPEKYPYLVKVSQEQTAQNAIWDGCIHAREWISPAFCFLWFI----- 173
Db 138 EMVSRIKIGSTVBDNPLVYVKI-GKDKGERKAIKFMDCGIIHAREWISPAFCQMFVYQATKS 196

Qy 174 -GHN-----NMWRKNRSFYANNHCIGTDLNRNFS 202
Db 197 YGKNKIMTKLLDRNMFVLPVFNVDGYIWSQTDNRWRKNRSNQNSTCIGTDLNRNF-D 255

Qy 203 KHWCEGSSSSCSSTYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYS 262
Db 256 VSWDSSPNTKPCLVNVRGPAPESEKETKAVTTFIRSHLSIKAYITHTSYSQMMLIPYG 315

Qy 263 YTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYS 322
Db 316 YTFKLPNQHQLLKVARIATDAL-STRYETRYIYGPASTYKTSGLSSLDWVYDLGKHT 374

Qy 323 F 323
Db 375 F 375

RESULT 4
A43929
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N/Alternate names: mast cell carboxypeptidase A3
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C/Accession: A43929; A39246; A45759
R/Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A/Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A.
A/Reference number: A43929; MUID:92105393; PMID:1729276
A/Accession: A43929
A/Molecule type: DNA
A/Residues: 1-417 <REY>
A/Cross-references: GB:M73716
A/Experimental source: mast cell
A/Note: The authors translated the codon CGC for residue 231 as Thr
R/Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Augstbaker, D.J.; Austen, K.F.; Serafin, W.E.
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A/Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison with the rat carboxypeptidase A.
A/Reference number: A39246; MUID:90083291; PMID:2594780
A/Accession: A39246
A/Molecule type: mRNA
A/Residues: 1-417 <RE2>
A/Cross-references: GB:M27717; NID:g179933; PIDN:AAA35652.1; PID:g179934
R/Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.U.

A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08254
A:Molecule type: protein
A;Residues: 16-40 <MOU>
A;Experimental source: pancreas
E:Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, P.X.
Eur. J. Biochem. 179, 609-616, 1989
A>Title: Purification and properties of five different forms of human procarboxypeptidase
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A:Molecule type: protein
A;Residues: 16-36,'X',38-43 <PAS>
A;Experimental source: pancreas
A;Accession: S02812
A:Molecule type: protein
A;Residues: 16-36,'Q',38-43 <PA2>
A;Experimental source: pancreas
C;Genetics:
A;Gene: GDB:CPBL
A;Cross-references: GDB:IJ2160; OMIM:114852
A;Map position: 3q21.3-q25
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-41/Domain: activation peptide #status predicted <ACP>
F;111-115/Product: carboxypeptidase B, pancreatic #status predicted <MAT>
F;173-186/Disulfide bonds: #status predicted
F;176,179,303/Binding site: zinc (His, Glu, His) #status predicted
F;355,377/Active site: Tyr, Glu #status predicted

Query Match 34.4%; Score 618; DB 1; Length 416;
Best Local Similarity 35.7%; Pred. No. 1.2e-43;
Matches 129; Conservative 68; Mismatches 120; Indels 44; Gaps 7;

Qy 1 PQSQVLAALPRTSRQVOVLNLTITYEIVLWQPVTADLIYKKKQHFFVNASDVNDVKR 60
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 22 FEGEKFRNVNEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDPRKAEDTVTEN 81

Qy 61 HLNVSIGPCSVLLADVDELIOOQISNDTSVPASASYEQYHSLSNEIYSWIPIFERHPD 120
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 82 VLKQNELQYKVILSNLRNVVAQPDSSRV---RATGSHVEKYNKWTIEAWTQQVATENPA 138

Qy 121 MLTKIHIGSSPEKPYLYLVKVGKGQTAKNAIWDCGHAREWISPAPCLMFT----- 173
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 139 LISRSVIGTTPEGRAIYLKLV-KRAGQNKPAIFDCGFHAREWISPAFCWFVRVAVTY 197

Qy 174 -----CH-----NRMRKNRSFYANNHCITGTDLNRNFASK 203
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 198 GREIQVTLELDKLDPYLPVLNIIDGYIYTWTKSFRWKTRSTHTGS-SIGTPNNRF--DA 255

Qy 204 HWCEGASSSCSETYCGLYPESEPEKVASFLRRNINQIKAYISMHSYSHVIPFPYSY 263
Db :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 256 GWCHIGASRNPCDETYCGPAASEKETKALADFIRNKUSSIKAYLTIHSYQMVIYPYSY 315

Qy 264 TRSKSKDHESLSVASEAVRAIEKTSKNTRYTHGSETLYLAPGGDDMIYDLGIKYSF 323
Db ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
 316 AYLGGENNAELNALAKATVKEL-ASHLGTKYTYPGATTIYPAAAGSGDDWAYDOGIRYSF 374

Qy 324 T 324
Db |
 375 T 375

RESULT 6
CPBOB
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence revision 24-Apr-1984 #text_change 02-Aug-1994
C;Accession: A93197; A92150; A00912
R;Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A>Title: Amino-acid sequence of bovine carboxypeptidase B.
A;Reference number: A93197; MUID:75217824; PMID:1057162

A;Accession: A93797
A;Molecule type: protein
A;Residues: 1-306 <TIT>
R;Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3758-3764, 1974
A;Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations of
A;Reference number: A92150; MUID:74260705; PMID:4833744
A;Accession: A92150
A;Molecule type: protein
A;Residues: 31-93;131-181;263-265;292-306 <SCH>
R;Schmidt, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A;Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A;Reference number: A92839; MUID:76265065; PMID:957425
A;Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R;Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A;Title: Isolation and sequence of peptides at the active center of bovine carboxypeptidase B
A;Reference number: A92051; MUID:70007159; PMID:5344132
A;Contents: annotation; active site
R;Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A;Title: Identification of a glutamic acid at the active center of bovine carboxypeptidase B
A;Reference number: A92119; MUID:73061487; PMID:4565668
A;Contents: annotation; active site
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;63-76,135-158,149-163/Disulfide bonds: #status experimental
F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F;246,268/Active site: Tyr, Glu #status experimental

Query Match 32.6%; Score 586; DB 1; Length 306;
Best Local Similarity 43.9%; Pred. No. 3.7e-41;
Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

Qy 98 YEYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPLVLYKVSKEQTAKNAIWDG 157
Db 6 YEKNNWETIEAWTEQVASENPDLISRSAGITTFGLNITLLKV-GRPGSKPAVFMDG 64

Qy 158 IHAREWISPAFLWFI-----GH-----NMWR 180
Db 65 FHAREWISPAFLWFI-----GH-----NMWR 124

Qy 181 KNRSPYANNHCIGTDLNRNPFASKHWCCEGSSSCSETYCGLYPESEPEVKAVASFLRN 240
Db 125 KTRSTRAGSCTGDLNRNPFASKHWCCEGSSSCSETYCGLYPESEPEVKAVASFLRN 183

Qy 241 INQIKAYISMHSYSQHVFPYSYTRSKSKOHELSLVAEAVRAIEKTSKNTRYTHGHGS 300
Db 184 LSSIKAYLTISYSQVMMLYPYSYDYKLPKNVVELTLAKGVKKL-ASLHGTTYSYGP 242

Qy 301 ETLYLAPGGDDWIYDLGIKYSFT 324
Db 243 TTIYPASGGDDWAYDQGIKYSFT 266

RESULT 7
A38395
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C;Accession: A38395; A33118
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A;Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme activation
A;Reference number: A38395; MUID:91105153; PMID:1988052
A;Accession: A38395
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-309 <COL>
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A;Reference number: A33118

A;Accession: A33118
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-309 <CO2>
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase
F;248,270/Active site: Tyr, Glu #status predicted

Query Match 30.7%; Score 552; DB 2; Length 309;
Best Local Similarity 42.6%; Pred. No. 2.5e-38;
Matches 112; Conservative 39; Mismatches 72; Indels 40; Gaps 5;

Qy 98 YEYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPLVLYKVSKEQTAKNAIWDG 157
Db 8 YAKYNDWKIVSTKQWKEKHPWVSRIKIGSTVEDNPLYVLI-GRKQGERKAIFMDG 66

Qy 158 IHAREWISPAFLWFI-----GH-----NMWR 180
Db 67 IHAREWISPAFLWFI-----GH-----NMWR 126

Qy 181 KNRSPYANNHCIGTDLNRNPFASKHWCCEGSSSCSETYCGLYPESEPEVKAVASFLRN 240
Db 127 KNRSPYANNHCIGTDLNRNPFASKHWCCEGSSSCSETYCGLYPESEPEVKAVASFLRN 185

Qy 241 INQIKAYISMHSYSQHVFPYSYTRSKSKOHELSLVAEAVRAIEKTSKNTRYTHGHGS 300
Db 186 LNSIKAYITFHSYSQVLLFPYGYTIKLPNHDLLKVARIATDVL-SSRYETRYIYGPIA 244

Qy 301 ETLYLAPGGDDWIYDLGIKYSF 323
Db 245 STIYKTSGLSDWAYDLGIKHTF 267

RESULT 8
A32128
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C;Accession: A32128
R;Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rut
J. Biol. Chem. 263, 17828-17836, 1988
A;Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and ev
A;Reference number: A32128; MUID:89034323; PMID:3182871
A;Accession: A32128
A;Molecule type: mRNA
A;Residues: 1-417 <GAR>
A;Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

Query Match 30.6%; Score 549.5; DB 2; Length 417;
Best Local Similarity 34.6%; Pred. No. 6.1e-38;
Matches 134; Conservative 63; Mismatches 127; Indels 63; Gaps 12;

Qy 1 FQSGQVLAALPRTSROVQLQLTTT--VEIVLWQPVTADLIYKKQVHFFVNASDVNV 58
Db 20 FVGQVLEIIPSHEEQRTLLQEAHEHLEDFWKSPT----IPGETVHVRFVFASTQAV 75

Qy 59 KAHLNVSIGPCSVLLADVEDLIQQQISN---DTVSPRASASYEYQYHSLNEIYSWIEFIT 115
Db 76 KVFLSQGDYSIMIEDVQVLLDQEREEMLFNQORERGGNFNFAYHTLEEIYQEMDNLV 135

Qy 116 ERHPDMLTKIHIGSSPEKYPLVLYKVS-GRKQAKNAIWDGCIHAREWISPAFLW--- 171
Db 136 AENPGLVSKVNLGSSFPENPMNLFKFTGDD---KPAIWLDAIGHAREWVQTALWTAN 192

Qy 172 -----FI-----GH-----NMWRKNRSFYANNHCIGTDLN 197
Db 193 KIAADYGTDPATISLNTLTDIFLPLVTNPDGYVFSQTTNRMWRKTRSKRSGSCGVDPN 252

Qy 198 RNFASKHWCCEGSSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHI 257

A;Contents: annotation; allelic variant
R;Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A;Reference number: A50963; PDB:4CFA
A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137, 'O', 139-141
R;Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 168, 367-387, 1983
A;Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A;Reference number: A92902; MUID:83294519; PMID:6887246
A;Contents: annotation; X-ray crystallography, 1.54 angstroms
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status experimental <ACP>
F;111-419/Product: carboxypeptidase A #status experimental <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F;248-271/Disulfide bonds: #status experimental
F;358,390/Active site: Tyr, Glu #status experimental

Query Match 29, 8%; Score 536.5; DB 1; Length 419;
Best Local Similarity 34.8%; Pred. No. 7.4e-37;
Matches 129; Conservative 59; Mismatches 124; Indels 59; Gaps 10;

```

Qy 1 FQSQVLAAALPRTSRQVQLQNLTITTYEIVL--W---QPVTADLIVKKQVHFFYNASD 54
Db 20 FVGHVLRITAADEAEVQTKVEDLEHLQLDFWRGPGQP-----GSPIDVRVPFPS 71
Qy 55 VDNVKAHLNVSGIPCSVLLADVEDLI---QQIISNDTVSPRASASY-YEQYHSLNEIYSW 110
Db 72 LQAVKVFLEAHGIRYIMIEDVQSLIDBEQEQMFASQSRARSTNTFNATYHTLOEIYDF 131
Qy 111 IEFITERHPDMLTKIHIGSSPEKPYLVYLVKSGKEQTAKNAIWIDCGITHAREWISPAFCL 170
Db 132 MLLVAEHPQLVSKLQIGRSGYEGRPYIVYVKFS-TGGSNRPALWIDLGHSHREWITQATGV 190
Qy 171 WFI-----GHNMRWRKNRSFYANNHCIG 193
Db 191 WFAKFTEDYQDDPSFTALLDSMDIFLEIVTNPDGFAFTHSQNRLWRKTRSVTSSSLCVG 250
Qy 194 TDLNRNPASKHWCBEAGSSSCSETYCGLYPESEPEVKAVASFLLRNINQIKAYISMHSY 253
Db 251 VDNARNW-DAGFGAGAGSSPCSETYHGKYANSEVEVKSIYDFVKDHGN-KFAFLSIHSY 308
Qy 254 SOHIVRPYSYTRSKDKHEELSLVASEAVRAIEKTSKNTRYTHGHGSETFLYLAPGGDDW 313
Db 309 SOLLLYPYGYTTSIPDKTELNQVAKSAVEAL-KSLYGTSTYKYSIITTIYQASGGSIDW 367
Qy 314 IYDLGIKYSFT 324
Db 368 SYNGIKYSFT 378

```

RESULT 10
CPRTA
carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.P.; Nikovits Jr., W.C.; Standring, D.N.; Pictet, R.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A;Molecule type: mRNA
A;Residues: 1-419 <QUI>
A;Cross-references: GB:J00713; NID:G203364; PIDN:AAA40893.1; PID:G203365
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparat
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A;Molecule type: DNA

A;Residues: 1-260, 'F', 262, 'M', 264-346, 'K', 348-419 <CLA>

C;Genetics:

A;Gene: CPA1

A;Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1

A;Note: the authors translated the codon GTC for residue 196 as Ile

C;Superfamily: carboxypeptidase

C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-110/Domain: activation peptide #status predicted <PRO>

F;111-419/Product: carboxypeptidase A #status predicted <MAT>

F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted

F;248-271/Disulfide bonds: #status predicted

F;358,380/Active site: Tyr, Glu #status predicted

Query Match 29.68; Score 532.5; DB 1; Length 419;

Best Local Similarity 35.3%; Pred. No. 1.6e-36;

Matches 131; Conservative 57; Mismatches 124; Indels 59; Gaps 10;

QY 1 FQSGVLAALPRTSRQVQLQLNTTYYEIVL--WQVPTADLIYKKQVHFFVNASVDVNV 58

DB 20 FVGHQVLRISADEAQVQKVELEHLEHLQDFWRDARAGIDIVRVFP-----PSIQSV 75

QY 59 KAHLNVSIGPCSVLLADVEDLI---QQISNDTVSPRASASY-YEQVHSLNEIYSWIEFI 114

DB 76 KAFLEVHGIGVEIMIEDVQLLDBEKKQMSAFQARALSTDSENFYATYHTLDEIVFMDLL 135

QY 115 TERHPDMLTKHTIGSSPEKPLYLVKVSKEQAKNAIWDGCIHAREWISPAFLWFIG 174

DB 136 VAEHPOLVSKIQIGNTPEGRPIHVLKFS--TCGTNRPAIWDGCIHAREWISPAFLWFIG 194

QY 175 H-----NRMWRKNSFYANNHCIGTDLN 197

DB 195 KVTYQYQDPTFTAVLDNMDIFLEIVNPDGPAYTHKTNRMWRKTRSHTSGLCVGVDPN 254

QY 198 RNFASKHW-----CEEGASSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSY 253

DB 255 RN-----WDAGLGKAGASSNPCSETYRGKFPNSEVEKSVIVDFVTSNGN-IKAFISHSY 308

QY 254 SQHIVPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTYTHGHGSETLYLAPGGGDDW 313

DB 309 SQLLYPYGYTSEPPAPQAEQLDLAKSAVTAL-TSLHGTFKYGSIIDTYQASGSDTDW 367

QY 314 IYDLGIKYSFT 324

DB 368 TYSQGIKYSFT 378

RESULT 11

A56171

carboxypeptidase A2 (EC 3.4.17.15) precursor - human

N;Alternate names: pancreatic carboxypeptidase A2

C;Species: Homo sapiens (man)

C;Date: 28-Apr-1995 #sequence revision 11-Aug-1995 #text_change 22-Jun-1999

C;Accession: A56171; S02809; S71395

R;Cataus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.

J. Biol. Chem. 270, 6651-6657, 1995

A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA

A;Reference number: A56171; MUID:95204457; PMID:7896805

A;Accession: A56171

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-417 <CAT>

A;Cross-references: CB:U19977; NID:G790226; PID:AAA74425.1; PID:G790227

A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340

R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.

Eur. J. Biochem. 179, 609-616, 1989

A;Title: Purification and properties of five different forms of human procarboxypeptidase

A;Reference number: S02809; MUID:89153096; PMID:2920728

A;Accession: S02809

A;Molecule type: protein

A;Residues: 'S', 18-36, 'N', 38-43 <PAS>

R;Laethem, R.M.; Blumenkopf, T.A.; Cory, C.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton

Arch. Biochem. Biophys. 332, 8-18, 1996

A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 ar

A;Reference number: S71394; MUID:96400327; PMID:8806703

A;Accession: S71395

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-301, 'T', 303-338, 'SR', 341-417 <LAE>

C;Genetics:

A;Gene: GDB:CPA2

A;Cross-references: GDB:125230; OMIM:600688

A;Map position: 7q32-7qter

C;Superfamily: carboxypeptidase

C;Keywords: hydrolase; metallo-carboxypeptidase; zinc

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-110/Domain: activation peptide #status predicted <ACP>

F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>

F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted

F;246-269/Disulfide bonds: #status predicted

F;356,378/Active site: Tyr, Glu #status predicted

Query Match 28.4%; Score 511.5; DB 2; Length 417;

Best Local Similarity 33.1%; Pred. No. 8.9e-35;

Matches 128; Conservative 61; Mismatches 135; Indels 63; Gaps 10;

QY 1 FQSGVLAALPRTSRQVQLQLNTTYYEIVL--WQVPTADLIYKKQVHFFVNASVDVNV 58

DB 20 FVGHQVLRISADEAQVQKVELEHLEHLQDFWRDARAGIDIVRVFP-----PSIQSV 75

QY 59 KAHLNVSIGPCSVLLADVEDLI---QQISNDTVSPRASASY-YEQVHSLNEIYSWIEFI 115

DB 76 KAFLEVHGIGVEIMIEDVQLLDBEKKQMSAFQARALSTDSENFYATYHTLDEIVFMDLL 135

QY 115 TERHPDMLTKHTIGSSPEKPLYLVKVSKEQAKNAIWDGCIHAREWISPAFLWFIG 174

DB 136 VAEHPOLVSKIQIGNTPEGRPIHVLKFS--TCGTNRPAIWDGCIHAREWISPAFLWFIG 194

QY 175 H-----NRMWRKNSFYANNHCIGTDLN 197

DB 195 KVTYQYQDPTFTAVLDNMDIFLEIVNPDGPAYTHKTNRMWRKTRSHTSGLCVGVDPN 252

QY 198 RNFASKHW-----CEEGASSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSY 257

DB 253 RN-----WDAGLGKAGASSNPCSDSYHGPSANSEVEKSVIVDFIKSH-GKVKAFILHSYSQL 310

QY 258 VRPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTYTHGHGSETLYLAPGGGDDW 317

DB 311 MFPIGYKTKLDFDELSEVAQAQSL-RSLHGTFKYGKVCISVIQASGSDWSYD 369

QY 318 GIKYSFTSN-----PPVEKLPL 334

DB 370 GIKYSFAFELRDTGRYGFLLPARQILP 396

RESULT 12

S29127

carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human

N;Alternate names: pancreatic carboxypeptidase A1

C;Species: Homo sapiens (man)

C;Date: 25-Feb-1994 #sequence revision 19-Jan-1996 #text_change 18-Jun-1999

C;Accession: S29127; A34205; S08253; S02810; S71394; S02811

R;Cataus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver,

Biochem. J. 287, 299-303, 1992

A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.

A;Reference number: S29127; MUID:93038569; PMID:1417781

A;Accession: S29127

A;Molecule type: mRNA

A;Residues: 1-419 <CAT>

A;Cross-references: EMBL:X67318; NID:G35329; PIDN:CAA47732.1; PID:G35330

R;Stewart, E.A.; Craik, C.S.; Hale, L.; Bowcock, A.M.

Am. J. Hum. Genet. 46, 795-800, 1990

A;Title: Human carboxypeptidase A identifies a BgIII RFLP and maps to 7q31-qter.

A;Reference number: A34205; MUID:90196012; PMID:1969228

A;Accession: A34205

A;Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA
A:Residues: 330-396 <STE>
A>Note: the authors translated the codon CTG for residue 391 as Val
R:Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A>Title: Further studies on the human pancreatic binary complexes involving procarboxypeptidase A
A:Reference number: S08253; MUID:90169111; PMID:2307232
A:Accession: S08253
A:Molecule type: protein
A:Residues: 17-43; 'XXX', 114-135 <MOU>
R:Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A>Title: Purification and properties of five different forms of human procarboxypeptidase A
A:Reference number: S02809; MUID:89153096; PMID:2920728
A:Accession: S02810
A:Molecule type: protein
A:Residues: 17-42 <PAS>
R:Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A>Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1
A:Reference number: S71394; MUID:96400327; PMID:8806703
A:Accession: S71394
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-419 <LAE>
C:Genetics:
A:Gene: GDB:CFAL; CPA
A:Cross-references: GDB:120597; OMIM:114850
A:Map position: 7q32-7qter
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-110/Domain: activation peptide #status predicted <ACP>
F:111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F:179-182,306/Binding site: zinc (His, Glu, His) #status predicted
F:248-271/Diulfide bonds: #status predicted
F:358,380/Active site: Tyr, Glu #status predicted

Query Match 27.4%; Score 493.5; DB 1; Length 419;
Best Local Similarity 33.1%; Pred. No. 2.8e-33;
Matches 126; Conservative 58; Mismatches 118; Indels 79; Gaps 12;

QY 1 PQSGQVLAALPRTSROVQVLONLTTTYEIVL--WQ-----PVTADLVKKKQVHFFVN 51
DB 20 FVGHQVLRISVADEAQVQKVELEDLEHLQLDQFMRGPAHGPSIDV-----RVPF--- 69

QY 52 ASDVDNVKAHLNVSGIPCSVLLADVEDLI--QQQI-----SNDTVSPRASASYEQ 100
DB 70 -PSIQAVKIFLESHGISYETWIEDVQSLDEQEQMFARFRARSTDTFN-----YAT 121

QY 101 YHSLNBIYSWIEPITERHPDMLTKIHGSSFEPKLYLVKVSKEQTAKNAIWDGCIHA 160
DB 122 YHTELEIYDFDLVAENPHLVSKIQIGNTYEGRPIYVLKFS-TGGSKRAIWDGTIHS 180

QY 161 REWISPAFCLWFI-----GHNRMWRKNR 183
DB 181 REWVTQASGVWFAKKITQDYGDAAFTALDITLDIFLEIVTNPDPGFATHTNRMWRKTR 240

QY 184 SFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRNINQ 243
DB 241 SHTAGSLCIGVDNRNWDAGFL-SSAGSNPCSETYHGKFPANSEVEKSIYDVPKDHGN- 298

QY 244 IKAYISMHSYSQHIVFPYPSYTRSKSDHELSIVASEAVRAIEKTSKNTRYTHGHGSETL 303
DB 299 IKAFISHSYQLLMYPYGYKTEPVPDQDLQSKAANTAL-ASLYGTFKFNYSIIKAI 357

QY 304 YLAPGGDDWYDLGIKYSFT 324
DB 358 YQASGGSTIDWYSGQIKYSFT 378

RESULT 13
T33527

hypothetical protein T06A4.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C:Accession: T33527
R:Wu, X.; Tin-Wollam, A.; Ozeresky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid T06A4.
A:Reference number: Z21364
A:Accession: T33527
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-528 <WUX>
A:Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A:Experimental source: strain Bristol N2; clone T06A4
C:Genetics:
A:Gene: CESP:T06A4.1
A:Map position: 1
A:Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C:Superfamily: carboxypeptidase

Query Match 24.6%; Score 443; DB 2; Length 528;
Best Local Similarity 32.6%; Pred. No. 6.1e-29;
Matches 121; Conservative 57; Mismatches 121; Indels 72; Gaps 14;

QY 11 PRTSROVQVLONL---TTTYEIVLWQPVLT-----ADLVKKKQVHFFVNASVDNVKAHL 62
DB 54 PETEGSVKYLRSLYEDPSPYELDFWQPTNIGAIVDLTVAPADAPRFV--KOLESKK--- 108

QY 63 NVSGIPCSVLLADVEDLIQQOISNDTVSPRASASYEOVHSLNEIYSWIEFITERHPDML 122
DB 109 ----ISYIVAVNDLSKAIENERGDSKFYNPVAGFAYDKYNSLSEIQTENKRLKKEYPTMI 164

QY 123 TKIHGSSFEPKLYLVKVSKEQT--AKNAIWDGCIHAREWISPAFCLWFIHG----- 175
DB 165 TLIDIGQSHENRTLLVMKIKGRNPLGSKISMMIDAGIHAREWIAPIATAM-YIAHELLLG 223

QY 176 -----NRWRKNRS-----FYANNHCIGTD 195
DB 224 YENDATVAKLMDHIDFYILVPMNPDPGYEYSREKNRMWRKNRSAPKCARQTFSTVCCSGVD 283

QY 196 LARNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQ 255
DB 284 LARNF--DNWFASGTSSSDCHDTYHGSAAFSEPSQAVRDQFLEQNTPE--AFISLSHSYSQ 340

QY 256 HIVFPYSYTR-SKSKD-HEELSIVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDW 313
DB 341 MWLIPYGRKQSPQDYHTGLRPLALRATKALVEL-YGTYQVGTGADLMYEASGGSHDW 399

QY 314 IY-DLGIKYSF 323
DB 400 AKGQLKVPYAY 410

RESULT 14
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T20507; T24289
R:Matthews, L.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19284
A:Accession: T20507
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-323 <WIL>
A:Cross-references: EMBL:Z78411; PIDN:CAB01647.1; GSPDB:GN00023; CESP:F02D8.4
A:Experimental source: clone F02D8
R:Wild, A.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19869
A:Accession: T24289
A>Status: preliminary; translated from GB/EMBL/DDBJ

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:03:36 ; Search time 13.4071 Seconds
(without alignments)
1185.568 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQL.....IKYSFTSNPPVEXLLPLSLK 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	691.5	38.5	415	1	P19223 rattus norv
2	637.5	35.5	417	1	P15086 homo sapien
3	635.5	35.3	401	1	P09955 sus scrofa
4	632	35.2	417	1	P15089 mus musculu
5	630.5	35.1	416	1	P55261 canis famil
6	622	34.6	417	1	P15088 homo sapien
7	586	32.6	306	1	P00732 bos taurus
8	552	30.7	309	1	P21961 rattus norv
9	549.5	30.6	417	1	P19222 rattus norv
10	536.5	29.8	419	1	P00730 bos taurus
11	533.5	29.7	419	1	P00731 rattus norv
12	514.5	28.6	417	1	P48052 homo sapien
13	493.5	27.4	419	1	P15085 homo sapien
14	491	27.3	421	1	Q9ui42 homo sapien
15	411.5	22.9	304	1	P42788 simulium vi
16	393	21.9	433	1	O02350 anopheles g
17	375.5	20.9	303	1	P04069 astacus flu
18	300	16.7	430	1	P38836 saccharomyc
19	286.5	15.9	424	1	P29068 thermocactin
20	283	15.7	451	1	P18143 streptomyc
21	259	14.4	434	1	P19041 streptomyc
22	122.5	6.8	1380	1	O75976 homo sapien
23	118.5	6.6	1377	1	O89001 mus musculu
24	116.5	6.5	707	1	ORC1_SCHPO
25	110.5	6.1	722	1	Q92100 mus musculu
26	109	6.1	734	1	Q96sm3 homo sapien
27	103.5	5.8	445	1	P70080 gallus gall
28	101	5.6	663	1	P79009 schizosacch
29	99	5.5	376	1	P54497 bacillus su
30	99	5.5	444	1	P17752 homo sapien
31	99	5.5	1255	1	Q15055 homo sapien
32	97	5.4	297	1	Y103_HUMAN
33	95	5.3	444	1	P17290 oryctolagus

34 94 5.2 1160 1 C8CA_BACTP
35 93 5.2 1451 1 SPT6_YEAST
36 92 5.1 454 1 CBPH_LOPAM
37 92 5.1 488 1 TY3H_ANGAN
38 91 5.1 262 1 YCJI_ECOLI
39 90.5 5.0 319 1 UN30_CABEL
40 90.5 5.0 434 1 CBPH_BOVIN
41 90.5 5.0 476 1 CBPH_RAT
42 90.5 5.0 3588 1 SRFI_BACSU
43 89.5 5.0 476 1 CBPH_MOUSE
44 89 4.9 481 1 TPH_XENLA
45 89 4.9 575 1 ACEB_YEAST

ALIGNMENTS

RESULT 1
CBPB_RAT
ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clausen E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
genes. Comparative analysis of the rat carboxypeptidase gene
family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-lysine (or L-arginine) + H(2)O =
peptide + L-lysine (or L-arginine).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M23959; AAA40872.1;
CC EMBL; M23947; AAA40872.1; JOINED.
CC EMBL; M23950; AAA40872.1; JOINED.
CC EMBL; M23952; AAA40872.1; JOINED.
CC EMBL; M23953; AAA40872.1; JOINED.
CC EMBL; M23954; AAA40872.1; JOINED.
CC PIR; A32129; A32129.
CC HSSP; P09955; INSA.
CC MEROPS; M14.003;
CC InterPro; IPR003146; Propep M14.
CC InterPro; IPR000834; Zn_carbopept.
CC Pfam; PF02244; Propep M14; 1.
CC Pfam; PF00246; Zn_carbopept; 1.
CC SMART; SM00631; Zn_pept; 1.
CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
CC Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
KW SIGNAL 1 13
FT CHAIN 14 108 ACTIVATION PEPTIDE.
FT PROPEP 109 415 CARBOXYPEPTIDASE B.
FT METAL 174 174 ZINC (BY SIMILARITY).
FT METAL 177 177 ZINC (BY SIMILARITY).
FT METAL 302 302 ZINC (BY SIMILARITY).


```
FT ACT SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
FT FT DISULFID 171 184 BY SIMILARITY.
FT FT DISULFID 243 266 BY SIMILARITY.
FT FT DISULFID 257 271 BY SIMILARITY.
SQ SEQUENCE 415 AA; 47515 MW; 9BA06CCADE30B6F2 CRC64;
Query Match 38.5%; Score 691.5; DB 1; Length 415;
Best Local Similarity 40.2%; Pred. No. 5.3e-52;
Matches 145; Conservative 59; Mismatches 114; Indels 43; Gaps 6;
Qy 1 FQSGQVLAALPRTSRQVQLQNTTVEIVLWQPVTDLVKKQVHFFVNASDVNVKA 60
Db 20 FDGNVYRVSVHGBDHVNLQELANTKEIDFWKPDSTQVKPLTTVDVHVKAEADV 79
Qy 61 HLNVSUPTPCSVLADVEDLQQQISNDTSPRASASYEQYHSLNEIYSWIFTERHPD 120
Db 80 FLEENEVHYEVLISNVNALESQFDSHT--RASGHSYTKYNKWETIEAWIQVATDNDP 136
Qy 121 MLTKIHIGSFPEKYPVLVVKSGEQAKNAIWDGCIHAREWISPAFCVLF-- 173
Db 137 LVTSQVIGTFEGNMYVLKI-GKTRNKPAIFIDCGFHAREWISPAFCVLFVREAVRTY 195
Qy 174 -----GH-----NRMWKNRSFYANNHCIGTDLNRNPAK 203
Db 196 NOBIHMKQLDELDFYVLPVNVINDGVYVWTQDMWRKTRSTWAGSSCLGVRENRF-NA 254
Qy 204 HWCEGASSSCSETYCGLPSEPEVKAVASFLRNINOIKAYISMHSYSHIVPYSY 263
Db 255 GWCEVGASRSCSETYCGPAPSEKETKALADFIKALYTHSYSQMWLYPYSY 314
Qy 264 TRSKDKHELSLVASAVRAIEKTSKNTRYTHGSGETLYLAPGGDDWIDYLGKYSF 323
Db 315 DYKLPENYEELNALVKAAKEL-ATLHGTYKTYGPGATTYPAGGSDDSWYQGIKYSF 373
Qy 324 T 324
Db 374 T 374
RESULT 2
ID -CPBP HUMAN STANDARD; PRT; 417 AA.
AC P15086; O60834; O96B08;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
DE (PASP)
DE CPB1 OR CPB OR PCPB.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC TISSUE=Pancreas;
RX MEDLINE=92129345; PubMed=1370825;
RA Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA French C.K.;
RT "Isolation of a cDNA encoding a human serum marker for acute
RT pancreatitis. Identification of a pancreas-specific protein as
RT pancreatic procarboxypeptidase B."
RL J. Biol. Chem. 267:2575-2581(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=98182241; PubMed=9524066;
RA Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J.,
RA Aviles F.X.;
RT "Comparative analysis of the sequences and three-dimensional models
RT of human procarboxypeptidases A1, A2 and B."
RL Biol. Chem. 379:149-155(1998).
RN [3]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalao D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 16-43.
RC TISSUE=Pancreas;
RX MEDLINE=89153036; PubMed=2920728;
RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT "Purification and properties of five different forms of human
RT procarboxypeptidases."
RL Eur. J. Biochem. 179:609-616(1989).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -I- TISSUE SPECIFICITY: Pancreas.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; M61057; AAA66973.1; -.
DR GO; GO:0004180; P:carboxypeptidase activity; TAS.
DR EMBL; AJ224866; CAAL2163.1; -.
DR EMBL; BC015338; AAH15338.1; -.
DR PDB; 1KWM; 28-AUG-02.
DR MEROPS; M14_003; -.
DR Genew; HGNC:2239; CPB1.
DR MIM; 114852; -.
DR GO; GO:0004180; P:carboxypeptidase activity; TAS.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn Carboxypept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn Carboxypept; 1.
DR SMART; SM00631; Zn pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN 1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN 2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 110 CARBOXYPEPTIDASE B.
FT CHAIN 111 417 ZINC (BY SIMILARITY).
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
FT DISULFID 259 273 BY SIMILARITY.
FT CONFLICT 16 16 H -> A (IN REF. 1; AA SEQUENCE).
FT CONFLICT 17 17 H -> Q (IN REF. 1; AA SEQUENCE).
```

RN	[4]	STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RP		MEDLINE=91027767; PubMed=2223783;
RX		Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RA		"Sequence-specific 1H NMR assignments and determination of the
RT		secondary structure for the activation domain isolated from
RT		pancreatic procarboxypeptidase B.;"
RL		Biochemistry 29:7515-7522(1990).
RN	[5]	
RP		STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX		MEDLINE=91114693; PubMed=1989879;
RA		Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
RT		"The NMR structure of the activation domain isolated from porcine
RT		procarboxypeptidase B.;"
RL		EMBO J. 10:11-15(1991).
RN	[6]	
RP		STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX		MEDLINE=93044373; PubMed=1422143;
RA		Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.
RA		Huber R., Wuethrich K.;
RT		"Comparison of the NMR solution structure with the X-ray crystal
RT		structure of the activation domain from procarboxypeptidase B.;"
RL		J. Biomol. NMR 2:1-10(1992).
CC		-1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC		peptide + L-lysine(or L-arginine).
CC		-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC		-1- DATABASE: NAME=Worthington enzyme manual;
CC		WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR		PDB; LPBA; 31-OCT-93.
DR		PDB; INSA; 24-DEC-97.
DR		MEROPS; M14.003; -.
DR		InterPro; IPR003146; Propep M14.
DR		InterPro; IPR000834; Zn_catOept.
DR		Pfam; PF02244; Propep M14; 1.
DR		Pfam; PF00246; Zn_catOept; 1.
DR		PRINTS; PR00765; CRBOXYPTASEA.
DR		SMART; SM00631; Zn pept; 1.
DR		PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR		PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW		Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW		3D-structure.
FT	PROPEP	1 95 ACTIVATION PEPTIDE.
FT	CHAIN	96 401 CARBOXYPEPTIDASE B.
FT	DISULFID	158 171
FT	DISULFID	230 253
FT	DISULFID	244 258
FT	METAL	161 161
FT	METAL	164 164
FT	METAL	289 289
FT	ACT_SITE	341 341
FT	ACT_SITE	363 363
FT	TURN	8 9
FT	STRAND	11 17
FT	HELIX	20 32
FT	STRAND	36 39
FT	HELIX	43 45
FT	STRAND	50 56
FT	HELIX	58 60
FT	HELIX	61 69
FT	TURN	70 72
FT	STRAND	74 79
FT	HELIX	82 87
FT	TURN	88 88
FT	TURN	101 102
FT	HELIX	107 120
FT	TURN	122 124
FT	STRAND	125 132
FT	TURN	134 135
FT	STRAND	138 144
FT	TURN	149 150
FT	STRAND	153 157
FT	TURN	162 163
FT	HELIX	165 181

```

TURN 182 184
FT HELIX 186 194
FT STRAND 196 200
FT HELIX 205 213
FT TURN 214 214
FT TURN 216 217
FT STRAND 224 224
FT TURN 226 227
FT STRAND 231 231
FT TURN 235 237
FT TURN 243 244
FT STRAND 249 249
FT TURN 252 253
FT TURN 255 256
FT STRAND 257 257
FT TURN 262 263
FT HELIX 266 277
FT TURN 278 281
FT STRAND 282 289
FT STRAND 294 297
FT TURN 307 308
FT HELIX 309 327
FT STRAND 332 335
FT HELIX 336 339
FT HELIX 347 353
FT TURN 354 355
FT STRAND 359 363
FT TURN 371 372
FT HELIX 376 378
FT HELIX 379 398
FT TURN 399 401
SQ SEQUENCE 401 AA; 45713 MW; 53129AF159A26348 CRC64;

Query Match 35.3%; Score 635.5; DB 1; Length 401;
Best Local Similarity 37.4%; Pred. No. 3.3e-47;
Matches 135; Conservative 64; Mismatches 119; Indels 43; Gaps 6;

QY 1 FQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPTADLVKKQVHFFNAGSDVDNKA 60
DB 7 FEGEKVFRVNVEDNISELHELASTRQIDFWKPDSTQIKPHSTVDFRVKASDILAVED 66

QY 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEYQVHSLNEIYSWIEFITERHPD 120
DB 67 FLEQNEIQVEVLINLRSVLEAQFDSRC---RTTGHSEYKYNWETTEANTEQVTSKNPD 123

QY 121 MLTKIHGSSFEKPYLVLVKVSKEQTAKNAIWDGCIHAREWISPAFCLEWFI----- 173
DB 124 LISRSALGTTFDGDNVLLKV-GRPGSNKPAIFMDCGFHAREWISQAFQWFRDVRV 182

QY 174 -----GH-----NEMWKNRFSFYANNHCICGTDLNENFASK 203
DB 183 GYEAHMTPEFLDNLDFYVPLVNLIDGYITWTQNRMRKTRSTNAGSSCTGTDPRNF-NA 241

QY 204 HWCEGASSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHVFPYSY 263
DB 242 GWCTVGASVNPENETVCGSRAEKEKTKALADFIRNLSIKAYLTHISYQMLLYPYSY 301

QY 264 TRSKSDHEELSLVASAVRAIEKTSKNRYTHGHGSETLYLAPGGGDWYDLGIKYSF 323
DB 302 DYKLPENDAEINSLAKGAVKEL-ASLYGTSYSYSGSTTIYPAGGSDWAYNQGIKYSF 360

QY 324 T 324
DB 361 T 361

RESULT 4
ID CBPC_MOUSE STANDARD; PRT; 417 AA.
AC P15089;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Mus musculus (Mouse).
OC Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062123; PubMed=2584208;
RA Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA Serafin W.E.;
RT "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT novel member of the carboxypeptidase gene family.";
RL J. Biol. Chem. 264:20094-20099(1989).
CC -|- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -|- SUBCELLULAR LOCATION: Secretory granules.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DE EMBL; J05118; AAA37369.1; -.
DR PIR; A34487; A34487.
DR HSSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR MGD; MGI:88479; Cpa3.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
DR Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
KW SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
SQ SEQUENCE 417 AA; 48790 MW; A2B300A068D1BA6D CRC64;

Query Match 35.2%; Score 632; DB 1; Length 417;
Best Local Similarity 38.2%; Pred. No. 7e-47;
Matches 138; Conservative 57; Mismatches 122; Indels 44; Gaps 7;

QY 1 FQSGQVLAALPRTSRQVQLNLTITTYEIVLWQPTADLVKKQVHFFNAGSDVDNKA 60
DB 21 FDEKVFVKLVQNEKHSVLKNTLQSIELDFWTFDAIHDIANVTYDFRVSEKESQTIQS 80

QY 61 HLNVSIGPCSVLLADVEDLIQQQIS-NDTVSPRASASYEYQVHSLNEIYSWIEFITERHP 119
DB 81 TLEQKHINYEILLHDLQEEIEKQFDVKDEIAGRHS---YAKYNDWDKIVSWTEKMLEKHP 137

QY 120 DMLTKIHGSSFEKPYLVLVKVSKEQTAKNAIWDGCIHAREWISPAFCLEWFI----- 173
DB 138 EMVSRIRIKGSTVEDNPLVVLKI-GKKDGERKALFMDCGIHAREWISPAFCQWFFVQATKS 196

QY 174 -GN-----RMRKNRFSFYANNHCICGTDLNENFAS 202
DB 197 YGKNKIMTKLLDRMNFYVLPVFVNVDGYIWSWTQDRMRKNRNRNQNSTCIGTDLNENF-D 255

QY 203 KHWCEGASSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQHVFPYS 262
```

```

Db      256 VSDSSVPKPCLVNVRGAPSEKTKAVTFIRSHLSIKAYITFTHSYQMLLPYG 315
Qy      263 YTRSKSDKHELSLAVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGKYS 322
Db      316 YTFKLPNHDLLKVARIDAL-STRYTRYIYGVPIASTIYKTSGSSLDWVYDLGKHT 374
Qy      323 F 323
Db      375 F 375

RESULT 5
CBPC_CANFA
ID CBPC_CANFA STANDARD; PRT; 416 AA.
AC P5261;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE membrane associated protein) (ZAP47).
GN CPB1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Fukuoka S.-I.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =
CC peptide + L-lysine (or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D78348; BAA11366.1; -.
DR HSP; P09955; IPBA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Propep_M14; 1.
DR PRINTS; PR00765; Zn_carboxept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 110 416 CARBOXYPEPTIDASE B.
FT METAL 175 175 ZINC (BY SIMILARITY).
FT METAL 178 178 ZINC (BY SIMILARITY).
FT METAL 303 303 ZINC (BY SIMILARITY).
FT ACT_SITE 377 377 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 244 267 BY SIMILARITY.
FT DISULFID 258 272 BY SIMILARITY.
SQ SEQUENCE 416 AA; 47595 MW; DFE1992CD52F8AB4 CRC64;

Query Match 35.1%; Score 630.5; DB 1; Length 416;
Best Local Similarity 36.6%; Pred. No. 9.4e-47;
Matches 132; Conservative 64; Mismatches 122; Indels 43; Gaps 6;

Qy      1 FQSGVLALPRTSRQVQLNLTITTEIVLVQPTADLVKKQVHFFNADSDNVKA 60
Db      21 FEKGFRVNVEDENHINLLHTLASTQIDFWKPDPSVTQIKPHSTADFRVKAEDILTVD 80

```

```

Qy      61 HLNVSIPGCVSLADVEDLIQQOISNDTVSPRASASYEYQYHSLNEIYSWIEFITERHPD 120
Db      81 FLQNELHYEVLINLRVLEQGRQV---PATGHSYKYNRWETIEAMTQOVTSENPD 137
Qy      121 MLTKIHIGSSFEKYPYLVKVGSKQETAKNAIWDGCIHAREWISPAFCLWFI----- 173
Db      138 LISRRSIGTTFEGRTIYLLKV-GRAGQNKPAIFMDCGFHAREWISPAFQWQFVREXIRTY 196
Qy      174 -----GH-----NRMRKNRSPFYANNHCICIGTDLNRNPFASK 203
Db      197 GQEIHWTELLDKLDFYVLPVGNIDGYVYTWKRWKRTSTQVGTGTCVGTDTNRF-DA 255
Qy      204 HMCBEGASSSCSETYCGLYPSEPEVKAVASPLRRNINOIKAYISMHSYSQHIVPYSY 263
Db      256 GWCKIGASRNPCDDETYCGPAAESEKETKALANFIRSNLSIKAYLTHSYSQMMLYPYSY 315
Qy      264 TRSKSKDHELSLAVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGKYSF 323
Db      316 DYKLTENNAELNALAKATVKEL-ATLHGKTYTYGPGATTIYPAGGSDDDWAYDQGIKYSF 374
Qy      324 T 324
Db      375 T 375

RESULT 6
CBPC_HUMAN
ID CBPC_HUMAN STANDARD; PRT; 417 AA.
AC P15038; Q95E94;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE (Carboxypeptidase A3).
GN CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=90083291; PubMed=2594780;
RA Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA Austen K.F., Serafin W.E.;
RT "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT comparison of the protein with mouse mast cell carboxypeptidase A and
RT rat pancreatic carboxypeptidases."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mast cells;
RX MEDLINE=92105393; PubMed=1729276;
RA Reynolds D.S., Gurley D.S., Austen K.F.;
RT "Cloning and characterization of the novel gene for mast cell
RT carboxypeptidase A."
RL J. Clin. Invest. 89:273-282(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zdobych B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```

FAHEY J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[4]
SEQUENCE OF 110-417 FROM N.A.
MEDLINE=92333165; PubMed=1629626;
Natsuka M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuka M.,
Wintroub B.U., Rutter W.J., Goldstein S.M.;
"Human skin mast cell carboxypeptidase: functional characterization,
cDNA cloning, and genealogy";
J. Invest. Dermatol. 99:138-145 (1992).
[5]
SEQUENCE OF 110-137.
MEDLINE=89214692; PubMed=2708524;
Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
"Human mast cell carboxypeptidase. Purification and
characterization";
J. Clin. Invest. 83:1630-1636 (1989).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
amino acid.
CC -I- SUBCELLULAR LOCATION: Secretory granules.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M2717; AAA35652.1; -
CC EMBL; M73720; AAA35658.1; -
CC EMBL; M73716; AAA35658.1; JOINED.
CC EMBL; M73717; AAA35658.1; JOINED.
CC EMBL; M73718; AAA35658.1; JOINED.
CC EMBL; M73719; AAA35658.1; JOINED.
CC EMBL; BC012613; AAH12613.1; -
CC EMBL; S40234; AAB22578.2; ALT_SEQ.
CC PIR; A43929; A43929.
CC HSP; P09955; INSA.
CC MEROPS; M14.010; -
CC Genew; HGNC:2298; CPA3.
CC MIM; 114851; -
CC GO; GO:003041; C:secretory granule; NAS.
CC GO; GO:0004182; F:carboxypeptidase A activity; TAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
CC InterPro; IPR003146; Propep M14.
CC InterPro; IPR000834; Zn_carbopept.
CC Pfam; PF02244; Propep M14; 1.
CC Pfam; PF00246; Zn_carbopept; 1.
CC SMART; SM00631; Zn_pept; 1.
CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT METAL 378 378 NUCLEOPHILE (BY SIMILARITY).
FT ACT SITE 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
FT CONFLICT 63 63 N -> K (IN REF. 3).
SQ SEQUENCE 417 AA; 48700 MW; 8CB90DB758117B24 CRC64;

Query Match 34.6%; Score 622; DB 1; Length 417;
Best Local Similarity 35.5%; Pred. No. 5.1e-46;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;
QY 1 FOSGQVLAALPRTSRQVQLNLTITTTVEIVLWQPVTDALIVKCKQVHFFVNASDVNVKA 60
DB 21 FDREKVRVQPKQKQADIKDLAKTNELDFWYPGATHVAANMVDVRSKESQAIQS 80
QY 61 HLNVSIPCSVLLADVEDLIQQQIS-NDTVSPRASAYEYQHSLEINLSEYWIEFITERHP 119
DB 81 ALDONKMYEILIHDLQEEIEKQFDYKEDIPGRHS---YAKYNNWKEIVAWTEKMDMDKYP 137
QY 120 DMLTKIHIGSSFEKYPYLYLVKSGKQOTAKNAIWDGCIHAREWISPAFLCWF----- 173
DB 138 EMVSRIKIGSTVEDNPLYVLIKI-GEKNERKALFMDCGIHAREWISPAFLCWFYQATKT 196
QY 174 -----GH-----NRMWRKNSFYANNHCITGTDLNRNFS 202
DB 197 YGRNKIMTKLLDRMNFVILPVFNVVDGYIWSWTKRMWRKNSKQNSKCICTDLNRNF-N 255
QY 203 KHWCEGASSSSCSEYCYGLYPESEPEVKAVASFLRNINQIKAYISMHSYSOHIVPPYS 262
DB 256 ASWNSIPNTNDPCADNYRGSAPSEKETKAVTNFIRSHLNEIKVYITFHSYSQWLLFPYG 315
QY 263 YTRSKSKDHEELSLVASEAVRALEKTSKNTRYTHGSETLYLAPGGGDDWIYDLGIKYS 322
DB 316 YTSKLPNHDIAKAKIGTDVL-STRYETRIYIGPIESTIYPISSSLDWAYDLGIKHT 374
QY 323 F 323
DB 375 F 375
RESULT 7
CBPB BOVIN STANDARD; PRT; 306 AA.
ID CBPB BOVIN AC P00732;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN CPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670 (1975).
RN [2]
RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RX MEDLINE=74260705; PubMed=4833744;
RA Schmidt J.J., Hirs C.H.W.;
RT "Primary structure of bovine carboxypeptidase B. Inferences from the
RT locations of the half-cystines and identification of the active site
RT arginine.";
RL J. Biol. Chem. 249:3756-3764 (1974).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX MEDLINE=76265065; PubMed=957425;
RA Schmid M.F., Herriott J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RL J. Mol. Biol. 103:175-190 (1976).
RN [4]
RP ACTIVE SITE.
RX MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
RT carboxypeptidase B.";

J. Biol. Chem. 244:5246-5253 (1969).
[5]
RN ACTIVE SITE.
RP MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RX "Identification of a glutamic acid at the active center of bovine
RT carboxypeptidase B.";
CC J. Biol. Chem. 247:7864-7869 (1972).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-lysine (or L-arginine) + H(2)O =
CC peptide + L-lysine (or L-arginine).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A93797; CPBOB.
DR POB; 1CPB; 30-SEP-83.
DR MEROPS; M14.003; -.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT METAL 135 158
FT DISULFID 149 163
FT METAL 66 66 ZINC.
FT METAL 69 69 ZINC.
FT METAL 194 194 ZINC.
FT ACT_SITE 246 246 NUCLEOPHILE.
FT ACT_SITE 268 268
SQ SEQUENCE 306 AA; C329D2655C44A172 CRC64;

Query Match 32.68; Score 586; DB 1; Length 306;
Best Local Similarity 43.94; Pred. No. 4.2e-43;
Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

QY 98 YEQVHSLNEIYWFIEFITERHPDMLTKIHIGSGFEKYPVLYLVKVSKEQTAKNAIWDG 157
DB 6 YEKNNWETIEATEQVASENPDISRSALGTFTFLGNTIYLLKV-GKPGSNKPAVFMDCG 64
QY 158 IHAREWISPAFCWFI-----GH-----NRMWR 180
DB 65 FHAREWISPAFCWFVFAVRVYGREITHMTEFLDKLDFFVLPVNVNIDGIYITVTRMWR 124
QY 181 KNSRFVANNHCIGTDLRNFASKHWCCEGASSSCSETYCYLPESPEPEVKAVASFLRN 240
DB 125 KTRSTRAGSCTGTDLRNF-DAGWCISGASNNPCSETYCGSAESESKEKAVADFIH 183
QY 241 INQIKAYISMHSYSHIVFPVYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 300
DB 184 LSSIKAYLTHSYSQMLFPYGYTKLPNNVELNTLAKGAVKKL-ASLHGTTYSYGP 242
QY 301 ETLYLAPGGDDWIYDLGIKYSFT 324
DB 243 TTIYPASGGDDWAYDQGIKYSFT 266

RESULT 8
CBPC_RAT
ID -CBPC_RAT STANDARD; PRT; 309 AA.
AC P21961;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,

RA Neurath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules.";
RL Biochemistry 30:648-655 (1991).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A38395; A38395.
DR HSP; P09955; INSA.
DR MEROPS; M14.010; -.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 68 71
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 196 196 ZINC (BY SIMILARITY).
FT ACT_SITE 248 248 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 270 270 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 65 78 BY SIMILARITY.
FT DISULFID 137 160 BY SIMILARITY.
FT VARIANT 1 1 MISSING (IN THE MAJOR FORM).
SQ SEQUENCE 309 AA; 35786 MW; 20330FABC3EE83EF CRC64;

Query Match 30.7%; Score 552; DB 1; Length 309;
Best Local Similarity 42.6%; Pred. No. 3.6e-40;
Matches 112; Conservative 39; Mismatches 72; Indels 40; Gaps 5;

QY 98 YEQVHSLNEIYWFIEFITERHPDMLTKIHIGSGFEKYPVLYLVKVSKEQTAKNAIWDG 157
DB 8 YAKYNDWNKIVSTKQWKEKHEPMWSRIKIGSTVEDNPLYVLKI-GRKDGKAKIFMDCG 66
QY 158 IHAREWISPAFCWFI-----GH-----NRMWR 180
DB 67 IHAREWISPAFCWFVFAVRVYQAQKSYGKNIMTKLLDRMNFVLPVNVNIDGIYITVTRMWR 126
QY 181 KNSRFVANNHCIGTDLRNFASKHWCCEGASSSCSETYCYLPESPEPEVKAVASFLRN 240
DB 127 KNSKNSPSTCTGTDLRNF-DVSWDSSPTNDPCLSVYRGPAPESEKETKAVTNFIRSH 185
QY 241 INQIKAYISMHSYSHIVFPVYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 300
DB 186 LNSIKAYLTHSYSQMLFPYGYTKLPNNHQLLKAVARIATDVL-SSRYETRIYGP 244
QY 301 ETLYLAPGGDDWIYDLGIKYSF 323
DB 245 STIYKTSGLSDWAYDGLGIKHTF 267

RESULT 9
CBPC_RAT
ID -CBPC_RAT STANDARD; PRT; 417 AA.
AC P1922;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=89034323; PubMed=3182871;
RA Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
RA Graf M., Rutter W.J.;
RT "A novel rat carboxypeptidase, CPA2: characterization, molecular

RX MEDLINE=83294519; PubMed=6887246;
RA Rees D.C., Lewis M., Lipscomb W.N.;
RT "Refined crystal structure of carboxypeptidase A at 1.54-A
RL J. Mol. Biol. 168:367-387(1983).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX MEDLINE=96003618; PubMed=7556081;
RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT "The three-dimensional structure of the native ternary complex of
RT bovine pancreatic procarboxypeptidase A with proproteinase E and
RT chymotrypsinogen C".
RL EMBO J. 14:4387-4394(1995).
RN [8]
RP VARIANT ALLELIC.
RX MEDLINE=69283620; PubMed=5817619;
RA Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Identification of the amino acid replacements characterizing the
RT allotypic forms of bovine carboxypeptidase A".
RL Biochemistry 8:2762-2768(1969).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -I- SUBUNIT: MONOMER. THE ZMOGEN IS SECRETED AS A TERNARY COMPLEX
CC COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC PROTEINASE E.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -I- DATABASE: NAME=worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M61851; AAA30426.1; -;
DR EMBL; M61851; AAA30427.1; -;
DR EMBL; Z33906; CAA83955.1; -;
DR PIR; JN0126; CPBOA.
DR PDB; 3CPA; 15-JAN-87.
DR PDB; 4CPA; 22-OCT-84.
DR PDB; 5CPA; 15-JAN-87.
DR PDB; 6CPA; 15-OCT-91.
DR PDB; 7CPA; 31-JAN-94.
DR PDB; 8CPA; 31-JAN-94.
DR PDB; 1CBX; 31-JAN-94.
DR PDB; 1CFS; 15-OCT-94.
DR PDB; 2CTB; 31-JAN-94.
DR PDB; 2CTC; 31-JAN-94.
DR PDB; 1ARL; 01-AUG-96.
DR PDB; 1ARM; 17-AUG-96.
DR PDB; 1BAV; 01-APR-97.
DR PDB; 1YME; 12-FEB-97.
DR PDB; 1CPX; 05-AUG-98.
DR PDB; 1PYT; 27-JAN-97.
DR PDB; 1ELL; 26-JUN-02.
DR PDB; 1ELM; 26-JUN-02.
DR PDB; 1F57; 06-SEP-00.
DR PDB; 1HDO; 15-NOV-01.
DR PDB; 1HBU; 15-NOV-01.
DR PDB; 1HEE; 23-NOV-01.
DR PDB; 1M4L; 04-MAR-03.
DR MEROPS; M14.001; -;
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW 3D-structure; Polymorphism.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A.
FT METAL 179 179 ZINC.
FT METAL 182 182 ZINC.
FT METAL 306 306 ZINC.
FT ACT SITE 358 358 PROTON DONOR.
FT ACT SITE 380 380 NUCLEOPHILE.
FT DISULFID 248 271
FT VARIANT 289 289 I -> V (IN ALLELIC VARIANT).
FT VARIANT 338 338 E -> A (IN ALLELIC VARIANT).
FT VARIANT 415 415 L -> V (IN ALLELIC VARIANT).
FT CONFLICT 95 95 S -> L (IN REF. 5).
FT CONFLICT 199 199 D -> N (IN REF. 3).
FT CONFLICT 203 203 D -> N (IN REF. 3).
FT CONFLICT 224 224 D -> N (IN REF. 3).
FT CONFLICT 232 232 Q -> E (IN REF. 3).
FT CONFLICT 295 295 D -> N (IN REF. 3).
FT TURN 114 116
FT TURN 119 120
FT TURN 125 138
FT HELIX 125 138
FT TURN 140 142
FT STRAND 143 150
FT TURN 152 153
FT STRAND 156 162
FT STRAND 171 176
FT TURN 180 181
FT HELIX 183 199
FT TURN 200 202
FT HELIX 204 212
FT STRAND 214 231
FT HELIX 223 231
FT TURN 232 232
FT TURN 234 235
FT STRAND 242 242
FT TURN 244 245
FT STRAND 249 249
FT TURN 253 254
FT TURN 261 262
FT STRAND 267 267
FT TURN 270 271
FT TURN 273 274
FT STRAND 275 275
FT TURN 280 281
FT STRAND 284 296
FT STRAND 299 307
FT TURN 311 314
FT TURN 324 325
FT HELIX 326 344
FT STRAND 349 352
FT HELIX 353 356
FT TURN 357 357
FT HELIX 364 370
FT TURN 371 372
FT STRAND 375 381
FT TURN 388 389
FT HELIX 393 395
FT HELIX 396 416
SQ SEQUENCE 419 AA; 47082 MW; 21B86407B3BFC452 CRC64;
Query Match 29.8%; Score 536.5; DB 1; Length 419;
Best Local Similarity 34.8%; Pred. No. 1.2e-38;
Matches 129; Conservative 59; Mismatches 124; Indels 59; Gaps 10;
QY 1 FOSGQVLAALPRTSRQVQLNLTTTYEIVL--W-----QFVTADLVKKQVHFFVNASD 54
DB 20 FVGHQVLRITAADEAEVQVKELEHQLQDFWRGQGP-----GSPIDVRVPPPS 71
QY 55 VDNVKAHLNVSGIPCSVLADVEDLI---QQQISNDTVSPRASAY-YQQYHSLNEIYSW 110
DB 72 LQAVKVFLEAHGIRYRMIEDVQSLDDEQEOMFASQSRARSTNTFNATYHTLTDEIYDF 131

CC TISSUE=Pancreas;
 RX MEDLINE=95204457; PubMed=7896805;
 RA Catus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
 RA Billeter M.;
 RT "The sequence and conformation of human pancreatic
 RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and
 RT three-dimensional model.";
 RL J. Biol. Chem. 270:6651-6657 (1995).
 RP SEQUENCE FROM N.A.

CC TISSUE=Brain, and Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Schenken C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg K.H., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]

CC X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=98111000; PubMed=9450539;
 RA Reverter D., Garcia-Saez I., Catus L., Vendrell J., Coll M.,
 RA Aviles F.X.;
 RT "Characterisation and preliminary X-ray diffraction analysis of human
 RT pancreatic procarboxypeptidase A2.";
 RL FEBS Lett. 420:7-10 (1997).
 RN [4]

CC X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
 RX MEDLINE=98046021; PubMed=9384570;
 RA Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
 RT "The three-dimensional structure of human procarboxypeptidase A2.
 RT Deciphering the basis of the inhibition, activation and intrinsic
 RT activity of the zymogen.";
 RL EMBO J. 16:6906-6913 (1997).
 CC -I- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
 CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
 CC residues.
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U19977; AAA74425.1; -;
 DR EMBL; BC007009; AAH07009.1; -;
 DR EMBL; BC014571; AAH14571.1; -;
 DR EMBL; BC015140; AAH15140.1; -;
 DR F01; A56171; A56171.
 DR PDB; 1AYE; 13-JAN-99.
 DR PDB; 1D7D; 12-JUL-00.
 DR PDB; 1O6X; 30-JAN-03.
 DR MEROPS; M14.002; -;
 DR Genew; HGNC:2297; CPA2.
 DR MIM; 60688; -;
 DR GO; GO:0004182; F:carboxypeptidase A activity; TAS.

DR GO; GO:0007039; P:vacuolar protein catabolism; TAS.
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR008334; Zn_carboxypeptidase.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF02246; Zn_carboxypeptidase.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR SMART; SM00631; Zn_pept; 1.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
 KW 3D-structure.
 FT SIGNAL 1 16 POTENTIAL.
 FT PROPEP 17 112 ACTIVATION PEPTIDE.
 FT CHAIN 113 417 CARBOXYPEPTIDASE A2.
 FT METAL 177 177 ZINC.
 FT METAL 180 180 ZINC.
 FT METAL 304 304 ZINC.
 FT ACT SITE 378 378 NUCLEOPHILE.
 FT DISULFID 246 269
 FT DISULFID 318 352
 FT CONFLICT 80 80 E -> G (IN REF. 2; AAH07009).
 FT CONFLICT 302 302 T -> I (IN REF. 1).
 SQ SEQUENCE 417 AA; 46828 MW; 3D9C073A98700A83 CRC64;
 Query Match 28.6%; Score 514.5; DB 1; Length 417;
 Best Local Similarity 33.1%; Pred. No. 9.2e-37;
 Matches 128; Conservative 62; Mismatches 134; Indels 63; Gaps 10;
 QY 1 FQSGQVLAALPRTSRQVQVQLNLTITTYEIVL--WQPTADLIYKKQVHFVNASDVNV 58
 DB 20 FVGDQVLEIVPSNEEQIKLLQLEAQEHQLDFWKSPT-----TPGETAHRVRVPVNVQAV 75
 QY 59 KAHLNVSGIPCSVLLADVEDLIQQISNDTIVSPRASAS---YVEQVHSLNEIYSWIEFIT 115
 DB 76 KYFLESQGIAYSIMIEDVQLDKENEMLFNRRERSGNFNFGAYHTLEEIQEMDNLV 135
 QY 116 ERHPDMLTKIHGSSPEKPYLYLVKVS-GKEQTAKNAIWDICGHAREWISPAFLWIFIG 174
 DB 136 AEHPGLVSKVNIIGSSPENRPMNVLFKSTGDD---KPAIMLDAGIHAREWVQTALWTAN 192
 QY 175 -----HNRMRKNSFYANNHCIGTDLN 197
 DB 193 KIVSDYKDPSTISILDALDIFLLPTNPDYVFSQTKNMRKRSKVSGLCVGVDPN 252
 QY 198 RNFASHKHWCEGASSSSCSETCYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHI 257
 DB 253 RHW-DAGFGGPGASNSPCSDSYHGPGANSEVEKSVDFIKSH-GKVKAFITLHSYSQLL 310
 QY 258 VPPYSYTRSKDKHBSLSLVAEAVRAIKTSKNTRYTHGHGSETLYLAPGGDDWYIDL 317
 DB 311 MFPYGYKCTKLDLDFDELSEVAQAQSL-RSLHGTQYKVGPIGCVIYQASGGSIDWSYDY 369
 QY 318 GIKYSFTSN-----PPVEKLIP 334
 DB 370 GIKYSFAFELDTGRYGFLLPAQILP 396
 RESULT 13
 CBPL_HUMAN
 ID CBPL_HUMAN STANDARD; PRT; 419 AA.
 AC P15085; Q9BS67;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
 GN CPA1 OR CPA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;

RESULT 15	CBPZ_SIMV1	STANDARD;	PRT;	304 AA.
AC	P42788;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).			
OS	Simulium vittatum (Black fly).			
OS	Eukaryota, Metazoa, Arthropoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidae;			
OC	Simuliidae; Simulium.			
OX	NCBI_TaxID=7192;			
[1]	SEQUENCE FROM N.A.			
RN	TISSUE=Gut;			
RC	MEDLINE=94093864; PubMed=8269093;			
RA	Ramos A., Mahowald A., Jacobs-Lorena M.;			
RT	"Gut-specific genes from the black fly Simulium vittatum encoding			
RT	trypsin-like and carboxypeptidase-like proteins.";			
RL	Insect Mol. Biol. 1:149-163(1993).			
CC	- FUNCTION: Involved in the digestion of the blood meal.			
CC	- SUBCELLULAR LOCATION: Secreted.			
CC	- TISSUE SPECIFICITY: Gut-specific.			
CC	- INDUCTION: By blood meal.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
EMBL; L08481; AAL18531.1; --				
DR	HSP; P48052; IAYE.			
DR	MEROPS: M14 UPA; --			
DR	InterPro: IPR000834; Zn carboxypept.			
DR	Pfam; PF00246; Zn_carboxypept; 1.			
DR	SMART; SM00631; Zn_pept; 1.			
DR	PROSITE; PS00132; CARBOXYPEPT ZN 1; 1.			
DR	PROSITE; PS00133; CARBOXYPEPT ZN 2; 1.			
KW	Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.			
FT	NON_TER 1			
FT	METAL 58 58 ZINC (BY SIMILARITY).			
FT	METAL 61 61 ZINC (BY SIMILARITY).			
FT	METAL 184 184 ZINC (BY SIMILARITY).			
FT	ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).			
FT	ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).			
FT	DISULFID 125 148 BY SIMILARITY.			
SO	SEQUENCE 304 AA; 34849 MW; 2E6E3FF9A6AA9144 CRC64;			
Query Match	22.9%; Score 411.5; DB 1; Length 304;			
Best Local Similarity	33.3%; Pred. No. 4.4e-28;			
Matches	94; Conservative 52; Mismatches 85; Indels 51; Gaps 9;			
Qy	100 QYHSLNBIYSWIBFITERHPDMLTKIHGSSFFKYPFLYVLKVGKQTAQNAIWDGHIH 159			
Db	1 QYHTLPIYSWLDRLVQEHPEHVEPVVVGKSYEGREIRGVKVSYKK--GNPVVWVSNIH 58			
Qy	160 AREWISPA---FCL-----WFI-----GH-----NMWRKNRS 184			
Db	59 AREWITAAATTTLYLNELLTSKNSTIREMAENYDWFYFPVTPDGYVYVTTDRMWRKTRS 118			
Qy	185 FYANNHCIGTDLARNFASKHWCEGASSSSCSSTYCGLYPESPEVKAVASFLRRNINQI 244			
Db	119 PNFDLSCAGTDPNRNW-NFHWMEQGGTSRRPCTTYGKKAFSEVETRSFSDFLTKLKGQI 177			
Qy	245 KAYISMHSYSQHIYFFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLY 304			

```
Qy      305 LAPGGGDWY--DLGIKYSFTSNP-----PVEKLIP 334
          |||   | | | : | : | : | : | : |
Db      237 PASGGSMDWAYDTLIDPIATYTYELPRDGNWGFLPANQIIP 278
```

Search completed: January 1, 2004, 19:29:00
Job time : 15.4071 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: January 1, 2004, 19:34:47 ; Search time 5219.6 Seconds
(without alignments)
2649.143 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQL.....IKYFTSNPPVEKLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-O=/cn2 1/USPTO spool/US09980881/runat_31122003_135915_26829/app.query.fasta_1.1628
-DB=GenEmbl -QPWT=fastap -SUPFIX=age -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CCN 1 1 8325 @runat_31122003_135915_26829 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1:	gb.ba.*	GenEmbl:*
2:	gb.htg.*	
3:	gb.in.*	
4:	gb.om.*	
5:	gb.ov.*	
6:	gb.pat.*	
7:	gb.ph.*	
8:	gb.pl.*	
9:	gb.pr.*	
10:	gb.ro.*	
11:	gb.sts.*	
12:	gb.sy.*	
13:	gb.un.*	
14:	gb.vi.*	
15:	em.ba.*	
16:	em.fun.*	
17:	em.hum.*	
18:	em.in.*	
19:	em.mu.*	
20:	em.om.*	
21:	em.or.*	
22:	em.ov.*	
23:	em.pat.*	
24:	em.ph.*	
25:	em.pl.*	
26:	em.ro.*	
27:	em.sts.*	
28:	em.un.*	

29:	em.vi.*	
30:	em.htg.hum.*	
31:	em.htg.inv.*	
32:	em.htg.other.*	
33:	em.htg.mus.*	
34:	em.htg.pln.*	
35:	em.htg.rod.*	
36:	em.htg.mam.*	
37:	em.htg.vrt.*	
38:	em.sy.*	
39:	em.htgo.hum.*	
40:	em.htgo.mus.*	
41:	em.htgo.other.*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1784	99.2	1546	9	AB011969	AB011969 Homo sapi
2	1749.5	97.3	1825	6	AR232183	AR232183 Sequence
3	1732	96.3	1272	6	AR086324	AR086324 Sequence
4	1732	96.3	1272	6	BD106653	BD106653 Method of
5	1732	96.3	1728	6	AX409750	AX409750 Sequence
6	1732	96.3	1728	6	AX706330	AX706330 Sequence
7	1732	96.3	1728	9	HUMPCPX	M75106 Human prepr
8	1732	96.3	1749	6	I16100	I16100 Sequence 2
9	1732	96.3	1749	6	I33526	I33526 Sequence 2
10	1727	96.1	1272	9	BT006936	BT006936 Homo sapi
11	1727	96.1	1272	12	BT007990	BT007990 Synthetic
12	1727	96.1	1715	9	BC007057	BC007057 Homo sapi
13	1494	83.1	1421	10	AB021968	AB021968 Mus muscu
14	1494	83.1	1490	10	AF164524	AF164524 Mus muscu
15	1484	82.5	1430	10	AF186188	AF186188 Mus muscu
16	1470	81.8	1425	10	AB042598	AB042598 Rattus no
17	697.5	38.8	1327	5	AF190274	AF190274 Bothrops
18	688	38.3	1441	5	AB099302	AB099302 Paralicht
19	646.5	36.0	1215	6	I67698	I67698 Sequence 1
20	646.5	36.0	1251	4	SSC133775	SSC133775 Sus scrof
21	637.5	35.5	1254	12	BT008111	BT008111 Synthetic
22	637.5	35.5	1263	6	A51896	A51896 Sequence 60
23	637.5	35.5	1263	6	A87530	A87530 Sequence 11
24	637.5	35.5	1263	6	AR085861	AR085861 Sequence 77
25	637.5	35.5	1284	6	A51913	A51913 Sequence 77
26	637.5	35.5	1284	6	AR085875	AR085875 Sequence
27	637.5	35.5	1310	9	BSA224866	AJ224866 Homo sapi
28	637.5	35.5	1454	9	BC015338	BC015338 Homo sapi
29	632	35.2	1442	10	MUSCARA	J05118 Mouse mast
30	630.5	35.1	1251	4	DOGZAP47	D78348 Canis famil
31	627	34.9	1427	10	RND67914	U67914 Rattus norv
32	627	34.9	1674	9	BC012613	BC012613 Homo sapi
33	622	34.6	1622	9	HUMCARC	M27717 Human mast
34	622	34.6	1622	11	G28614	G28614 human STS S
35	619.5	34.5	1302	6	AX616688	AX616688 Sequence
36	619.5	34.5	1311	6	AR225025	AR225025 Sequence
37	619.5	34.5	1311	6	AX106749	AX106749 Sequence
38	619.5	34.5	1993	6	AX704687	AX704687 Sequence
39	619.5	34.5	2154	6	A67368	A67368 Sequence 12
40	618	34.4	1332	6	AX333659	AX333659 Sequence
41	618	34.4	1332	9	HUMPAS	M81057 Human proca
42	617.5	34.3	1307	9	BC033684	BC033684 Homo sapi
43	610	33.9	927	6	AR072911	AR072911 Sequence
44	606.5	33.7	1672	5	BC041315	BC041315 Xenopus 1
45	587	32.6	1339	5	GGCBPAMR	X64539 G.domesticu

ALIGNMENTS

RESULT 1

JOURNAL Patent: US 6455294-A 1 24-SEP-2002;
FEATURES Location/Qualifiers
source 1..1625
/organism="unknown"
BASE COUNT 488 a 334 c 315 g 488 t
ORIGIN
Alignment Scores:
Pred. No.: 7 02e-164 Length: 1625
Score: 1749.50 Matches: 336
Percent Similarity: 94.38% Conservative: 0
Best Local Similarity: 94.38% Mismatches: 2
Query Match: 97.30% Indels: 18
DB: 6 Gaps: 1

US-09-980-881A-3 (1-338) x AR232183 (1-1625)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 83 TTTGAGAGTGGCCNAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 142

Qy 21 GlnAsnLeuThrThrThrThrGluLeuValLeuTrpGlnProValThrAlaAspLeuLe 40
Db 143 CAGAACTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATT 202

Qy 41 ValIysLysGlnValHisPhePheValAsnAlaSerAspValAspValLeuAla 60
Db 203 GTGAAGAAAAAACAAGTCCATTTTGTAAATGTCATCTGATGTGCAAAATGTGAAGCC 262

Qy 61 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
Db 263 CATTTAANTGTAGCGGAATTCATGACGATGTTCTGTCGACATGTGGAAGATCTTATT 322

Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 323 CAACAGCAGATTTCACAGACACAGTCAGCCCCGAGCCCTCCGATCGTACTATGAACAG 382

Qy 101 TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 383 TATCACTCACTAAATGAATCTATTCTTGGATGAAATTTATACTGAGAGGCATCTCTGAT 442

Qy 121 MetLeuThrIysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 443 ATGCTTACAAAATTCACATGGATCTCTCATTTGAGAAGTACCACCTCTATGTTTAAAG 502

Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 503 GTTCTGGAAGAAGACAGCAGCCAAAATGCCATATGGATTGCTGTGGAATCCATGCC 562

Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
Db 563 AGAAGATGGATCTCTCTGCTTCTGCTTGTGGTTTCATAGGCCAATATCGAATGTGAGA 622

Qy 181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
Db 623 AGAAGCGGTTCTTCTATGCGAACAATCATTTGATCGGAACAGACACCTGGAATGGAAC 682

Qy 201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 220
Db 683 GCTTCCAAACACTGGTGTGAGAGGTGATTCATGCTCTGCTCGGAACCTACTGT 742

Qy 221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
Db 743 GGACTTTATCCTGAGTCAAGAACAGAGTCAAGGCACTGCTGCTTCTTGAAGAAAT 802

Qy 241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
Db 803 ATCAACAGATTAAGCATACATCAGCATGCTTCTATCTCCAGCATATAGTGTTCCTCA 862

Qy 261 TyrSerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGlu 280
Db 863 TATTCCTATACAGAGTAAGCAAGACCATGAGGACTGCTCTCTAGTAGCAGGTGA 922

Qy 281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
Db 923 GCAGTTCGTGCTATTGAGAAAAATTAGTAAAAATACCAGGTATACACATGCCATGGCTCA 982

Qy 301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
Db 983 GAAACCTTATACCTAGCTCTCTGGAGGTGGGACGATGGATCTATGATTTGGGCATCAA 1042

Qy 321 TyrSerPhe----- 323
Db 1043 TATTCGTTTACAATTGAACCTTCAGATATCGGCACATACGGAATTCCTGCTGCCGAGCGT 1102

Qy 324 -ThrSerAsnProProValGluLysLeuProLeuSerLeuLys 338
Db 1103 TACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAAA 1148

RESULT 3
LOCUS AR086324 1272 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5985562.
ACCESSION AR086324
VERSION AR086324.1 GI:10013090
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Morser, M. John. and Nagashima, M.
TITLE Method of detecting thrombotic disease risk associated with plasma carboxypeptidase B polymorphisms
JOURNAL Patent: US 5985562-A 1 16-NOV-1999;
FEATURES Location/Qualifiers
source 1..1272
/organism="unknown"
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN

Alignment Scores:
Pred. No.: 2 78e-162 Length: 1272
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-3 (1-338) x AR086324 (1-1272)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 67 TTTGAGAGTGGCCNAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126

Qy 21 GlnAsnLeuThrThrThrThrGluLeuValLeuTrpGlnProValThrAlaAspLeuLe 40
Db 127 CAGAACTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATT 186

Qy 41 ValLysLysGlnValHisPhePheValAsnAlaSerAspValAspValLeuAla 60
Db 187 GTCAAGAAAAAACAAGTCCATTTTGTAAATGCATCTGATGTGCACAATGTGAAAGCC 246

Qy 61 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
Db 247 CATTTAAATGTGAGCGGAATTCATGCAAGTGTCTGTCGACAGCGTGAAGATCTTATT 306

Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 307 CAACAGCAGATTTCACAGACACAGTCAGCCCCGAGCCCTCCGATCGTACTATGAACAG 366

Qy 101 TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 367 TATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATACTGAGAGGCATCTCTGAT 426

Qy 121 MetLeuThrIysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Db 787 CACTGGTGTGAGGAAGTGTCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 846
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnGln 243
Db 847 CCTGAGTCAGAACAGAGTGAAGGAGTGGCTAGTTCTTTGAGAGAAATATCAACAG 906
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 907 ATTAAGCATACATCAGCATGCTATCTCCAGCATATAGTGTTCATATTCCTAT 966
Qy 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAlaValArg 283
Db 967 ACAGAAAGTAAAGCAAGACCATGAGCAATGCTCTAGTAGCAGTGAAGCAGTTCT 1026
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1027 GCTATTGAGAAAACCTAGTAAAAATACCAGGTATACATGGCCATGGCTCAGAAACCTTA 1086
Qy 304 TyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1087 TACCTAGCTCCTGAGGTGGGACGATTGGATCTATGATTGGGCATCAAAATATTCGTTT 1146
Qy 324 -----ThrSerAs 326
Db 1147 ACAATTGAACCTCGAGATACGGGCACATACGGATTCTTGCTGCGGAGCGTTACATCAA 1206
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1207 CCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1243

RESULT 5
AX409750
LOCUS AX409750 1728 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2397 from Patent WO0229103.
ACCESSION AX409750
VERSION AX409750.1 GI:21442455
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 2397 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source
1..1728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. M75106"

BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN

Alignment Scores:
Pred. No.: 4.16e-162 Length: 1728
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-3 (1-338) x AX409750 (1-1728)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProHrghThrSerArgGlnValGlnValLeu 20
Db 86 TTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145
Qy 21 GlnAsnLeuThrThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuLe 40
Db 146 CAGAACTTACTACAACTATGAGATTGTCTCTGGCGCGCGGTAAACGCTGACCTTATT 205

RESULT 6

Qy 41 ValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60
Db 206 GTGAAGAAAAAACAAGTCCATTTTTTGTAAATAGCATCTCATGTGCAATGTGAAGACC 265
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
Db 266 CATTAAATGTGAGCGGAATTCATGCAGTGTCTTGTGCGACAGCTGGGAAGATCTTATT 325
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerIleTyrTyrGluGln 100
Db 326 CAACAGCAGATTTCCAAACGACACAGTCAGCCCCGCGCATCCGCTACTATGAACAG 385
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArgHisProAsp 120
Db 386 TATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAACTGAGAGCGCATCTCGAT 445
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 446 ATGCTTACAAAATCCACATGGATCTCTATTGAGAACTACCACTCTATGTTTAAAG 505
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrIleAspCysGlyIleHisAla 160
Db 506 GTTCTCGAAAAAGAACAAACAGCAAAATGCCATATGGATTGACTGTGGAATCCATGCC 565
Qy 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis----- 175
Db 566 AGAAGATGGATCTCTCTGCTTTCTGCTTGTGTTTCATAGGCCATATAACTCAATTCAT 625
Qy 175 ----- 175
Db 626 GGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTCTTATGTTATGCCGGTG 685
Qy 176 -----AsnArgMetTyrArgLysAsnArg 183
Db 686 GTTAATGTGACGGTTATGACTACTCATGCAAAAGAAATCGAATGTGAGAAAGAACCGT 745
Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 746 TCTTTCTATCGCAACAATCATTCATCGCAACAGACCTGAATAGGAACCTTTGTCTCCAA 805
Qy 204 HisTyrCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 806 CACTGGTGTGAGGAAGTGATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTAT 865
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 866 CCTGAGTCAGAACCAAGAGTGAAGGAGTGGCTAGTTCTTGTGAGAGAAATATCAACAG 925
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 926 ATTAAGCATACATCAGCATGCTATTCATCTCCAGCATATAGTGTTCATATTCCTAT 985
Qy 264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 986 ACACGAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCAGTTCTGT 1045
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1046 GCTATTGAGAAAACCTAGTAAAAATACAGGTATACATGCGCCATGGCTCAGAAACCTTA 1105
Qy 304 TyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1106 TACCTAGCTCCTGGAGGTGGGACGATTGGATCTATGATTGGGCATCAAAATATTCGTTT 1165
Qy 324 -----ThrSerAs 326
Db 1166 ACAATTGAACCTCGAGATACGGGCACATACGGATTCTTGCTGCGGAGCGTTACATCAA 1225
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1226 CCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1262

```
AX706330
LOCUS AX706330 1728 bp DNA linear PAT 04-APR-2003
DEFINITION Sequence 1 from Patent WO03014391..
ACCESSION AX706330
VERSION AX706330.1 GI:29562754
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Morten, J.E. and Jackson, K.N.
TITLE Patent: WO 03014391-A 1 20-FEB-2003;
JOURNAL AstraZeneca AB (SE)
FEATURES
    source
        1..1728
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN
Alignment Scores:
Pred. No.: 4..16e-162 Length: 1728
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: Gaps: 2
US-09-980-881A-3 (1-338) x AX706330 (1-1728)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 86 TTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAGACCTCTAGGCAAGTTCAAGTTCTA 145
Qy 21 GlnAsnLeuThrThrThrGlyLeuValLeuTyrGlnProValThrAlaAspLeuLe 40
Db 146 CAGAAATCTTACTACCAATATGAGATTGTTCTCTGGCAGCGGTAACAGCTGACCTTATT 205
Qy 41 ValIleLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
Db 206 GTGAGAGAAAACAGTCCATTTTGTAAATGATCTGATGTCGACATGTGAAGCC 265
Qy 61 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
Db 266 CATTAAATGTGAGCGGAATTCATGCAGTGTCTTCTGGCAGACGTGGAAGATCTTATT 325
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 326 CAACAGCAGATTTTCCACGACACAGTCAAGCCCGGACCTCCGATCTGATATGAACAG 385
Qy 101 TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 386 TATCACTCACTAATGAATCTATTCTTGGATAGAAATTTAATCACTGAGAGCATCTCGAT 445
Qy 121 MetLeuThrIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 446 ATGCTTACAAAATCCACATTTGGATCTCTATTGAGAAGTACCACACTCTATGTTTAAAG 505
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 506 GTTTCGGAAGAAGCAAAACAGCCAAATGCCCATATGGATTTGATCTGGAATCCATGCC 565
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
Db 566 AGAAGATGGATCTCTCTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 625
Qy 175 ----- 175
Db 626 GGGATAATAGGGCAATATACCAATCTCTCGAGGCTTGTGGATTCTTATGTTATGCCGGTG 685
```

```
176 -----AsnArgMetTrpArgLysAsnArg 183
686 GTTAATGTGGACGGTTATGACTACTCATCGAAAAGAAATCGAATGTGGAGAAAGACCGT 745
184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
746 TCTTTCTATGCGAAACAATCATTCATCGGAACAGACCTGAATAGGAACCTTGTCTCCAAA 805
204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
806 CACTGTGTGAGGAAGTGCATCCAGTCTCTCATGCTCGAAACCTACTGTGGACTTTAT 865
224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
866 CTTGAGTCAGAACCAAGTGAAGGAGTGGCTAGTCTTCTGAGAAGAAATATCAACCCAG 925
244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
926 ATTAAGCATATACATCAGCATGCATTCTACTCCAGCATATAGTGTTCATATTCCTAT 985
264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
986 ACACGAAGTAAAGCAAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1045
284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
1046 GCTATTGAGAAAACCTAGTAAATACAGGTATACATGCGCCATCGCTCGAAACCTTA 1105
304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
1106 TACCTAGTCTCTGGAGTGGGAGCAGTGGATCTATGATTTGGGCATCAATATTCGTTT 1165
324 -----ThrSerAs 326
1166 ACAATTGAATCTCGAGATACGGGCACATAGCGATCTTCTGTCGGAGCGTTACATCAA 1225
326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
1226 CCCACCTGTAGAGAGCTTTTCCGCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1262
RESULT 7
HUMPCPBX 1728 bp mRNA linear PRI 07-JAN-1995
LOCUS Human prepro-plasma carboxypeptidase B mRNA, complete cds.
DEFINITION M75106
ACCESSION M75106.1 GI:189686
VERSION plasma carboxypeptidase.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Eaton, D.L., Malloy, B.E., Tsai, S.P., Hentzel, W. and Drayna, D.
TITLE Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
JOURNAL J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE 92042093
PUBMED 1939207
COMMENT Original source text: Homo sapiens liver cDNA to mRNA.
FEATURES
    source
        1..1728
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /tissue_type="liver"
                /gene="PCPB"
                /genes="PCPB"
                /notes="prepro-plasma carboxypeptidase B"
                /codon_start=1
                /protein_id="AAA60042.1"
                /db_xref="GI:189687"
```



```
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaSerValGluAspLeuLeu 80
Db 287 CATTTAAATGTGACGGAAATCCATGAGTGTCTTGTGGCAGACGTTGGAGATCTTATT 346

Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 347 CAACAGCAGATTTTCCACGACACAGTCCAGCCCGAGCCTCCGATCGTACTATGAACAG 406

Qy 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 407 TATCAGTCACTAAATGAAATCTATTCTTGGATAGATTTATACTGAGAGCATCTCGAT 466

Qy 121 MetLeuThrLysIleHisIleGlySerPheGluLysTyrProLeuTyrValLeuLys 140
Db 467 ATGCTTACAAAATCCACATTTGGATCTCTATTGAGAGGTACCCACTCTATGTTTAAAG 526

Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 527 GTTTCCTGGAAGAAGAAACAAACAGCCAAAATGCCATATGGATGGATCGTGAATCCATGCC 586

Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
Db 587 AGAGAATGGATCTCTCTGCTTTCTGCTTGTGGTTTCATAGGCCATATAAATCAATTCAT 646

Qy 175 ----- 175

Qy 647 GGGTAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy 176 -----AsnArgMetTrpArgLysAsnArg 183
Db 707 GTTAATGTGACGGTTATGACTACTCATGCAAGAAAGAAATCGAATGTGGAGAAAGACCGT 766

Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 767 TCTTTCTATCGAACAATCATTTGATCGAAGACACCTGATAGGAATTTGCTTCCAAA 826

Qy 204 HisTrpCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 827 CACTGGTGTGAGGAGGTGATTCAGTCTCTCATGCTCGGAACCTACTGTGACCTTTAT 886

Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 887 CCTGAGTCAGAACCCAGAAAGTGAAGGCAGTGGCTAGTTTCTTGAGAGAAATATCAACAG 946

Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 947 ATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTGTTTCCATATTCCTAT 1006

Qy 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAlaValArg 283
Db 1007 ACAGAAAGTAAAGAACCAATGAGAACTGCTCTGTAGTACCGAGTGAAGCATTCGT 1066

Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1067 GCTATTGAGAAACTAGTAAATAATACCAAGTATACATAGCCATGGCTCAGAAACCTTA 1126

Qy 304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1127 TACCTAGTCTCTGAGGTGGGACGATTTGATCTATGATTTGGGCATCAAAATATTCGTTT 1186

Qy 324 -----ThrSerAs 326
Db 1187 ACAATTGAATTCGAGATACGGGCACATACGGATTTCTGTCGGGAGCGTTATCATCAAA 1246

Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1247 CCCACCTGTAGAGAGCTTTTGCGCTGTCTCTAAAA 1283
```

RESULT 9

```
133526
LOCUS 133526 1749 bp DNA linear PAT 06-FEB-1997
DEFINITION Sequence 2 from patent US 5593674.
ACCESSION 133526
```

```
VERSION 133526.1 GI:1824317
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Plasma carboxypeptidase
JOURNAL Patent: US 5593674-A 2 14-JAN-1997;
FEATURES Location/Qualifiers
source 1..1749
/organism="unknown"
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN
```

Alignment Scores:

```
Pred. No.: 4,23e-162 Length: 1749
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 6 Gaps: 2
```

US-09-980-881A-3 (1-338) x 133526 (1-1749)

```
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 107 TTTTCAGATGGCCCAAGTTCTAGCTGCTCTTCTTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuLeu 40
Db 167 CAGAACTTACTACACATATGAGATTTCTCTGGCAGCCGTAACAGCTGACCTTATT 226

Qy 41 ValLysLysGlnValHisPhePheValAsnAlaSerAspValAspLeuAla 60
Db 227 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTCATGTGCACAAATGTGAAGCC 286

Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLeu 80
Db 287 CATTTAAATGTGAGCGGAATTCATGAGTGTCTTCTGGCAGCGTGGGAAGATCTTATT 346

Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 347 CAACAGCAGATTTTCCACGACACAGTCCAGCCCGAGCCTCCGATCGTACTATGAACAG 406

Qy 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 407 TATCAGTCACTAAATGAAATCTATTCTTGGATAGAAATTTATACTGAGAGCATCTCGAT 466

Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 467 ATGCTTACAAAATCCACATTTGGATCTCTCTATTGAGAGGTACCCACTCTATGTTTAAAG 526

Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 527 GTTTCCTGGAAGAAGAAACAAACAGCCAAAATGCCATATGGATGGATCGTGAATCCATGCC 586

Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
Db 587 AGAGAATGGATCTCTCTGCTTTCTGCTTGTGGTTTCATAGGCCATATAAATCAATTCAT 646

Qy 175 ----- 175

Db 647 GGGTAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy 176 -----AsnArgMetTrpArgLysAsnArg 183
Db 707 GTTAATGTGACGGTTATGACTACTCATGCAAGAAAGAAATCGAATGTGGAGAAAGACCGT 766

Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 767 TCTTTCTATCGAACAATCATTTGATCGAAGACACCTGATAGGAATTTGCTTCCAAA 826

Qy 204 HisTrpCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 827 CACTGGTGTGAGGAGGTGATTCAGTCTCTCATGCTCGGAACCTACTGTGACCTTTAT 886

Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 887 CCTGAGTCAGAACCCAGAAAGTGAAGGCAGTGGCTAGTTTCTTGAGAGAAATATCAACAG 946

Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 947 ATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTGTTTCCATATTCCTAT 1006

Qy 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAlaValArg 283
Db 1007 ACAGAAAGTAAAGAACCAATGAGAACTGCTCTGTAGTACCGAGTGAAGCATTCGT 1066

Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1067 GCTATTGAGAAACTAGTAAATAATACCAAGTATACATAGCCATGGCTCAGAAACCTTA 1126

Qy 304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1127 TACCTAGTCTCTGAGGTGGGACGATTTGATCTATGATTTGGGCATCAAAATATTCGTTT 1186

Qy 324 -----ThrSerAs 326
Db 1187 ACAATTGAATTCGAGATACGGGCACATACGGATTTCTGTCGGGAGCGTTATCATCAAA 1246

Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1247 CCCACCTGTAGAGAGCTTTTGCGCTGTCTCTAAAA 1283
```

```

Qy 204 HiiTrrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 827 CACTGGTGTGAGGAAGTGATCCAGTCTCTCATGCTCGAACCCTACTGTGGACTTTAT 886

Qy 224 ProGluSerGluProGluValValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 887 CCTGAGTCAGAACCAAGAGTGAAGGAGTGGCTAGTCTTCTTGAGAGAAATATCAACCAG 946

Qy 244 IleGlyAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 947 ATTAAGACATACATCAGCATGCATTCATCTCCAGCATATAGTGTTCATATTCCTAT 1006

Qy 264 ThrArgSerIlySerIlyAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 1007 ACAGGAAGTAAAGCAAGACCAAGAGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy 284 AlaIleGluIlyThrSerIlyAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1067 GCTATTGAGAAACTAGTAAATACCAAGTATACATGCGCATGGCTCAGAAACCTTA 1126

Qy 304 TyrLeuAlaProGlyGlyGlyAspTyrIleTyrAspLeuGlyIleTyrSerPhe 323
Db 1127 TACTAGTCTCTGAGGTGGGACGATTGGATCTATGATTGGGCATCAATATTCGTTT 1186

Qy 324 -----ThrSerAs 326
Db 1187 ACAATTGAATCTCGAGATACGGGCACATACGGAATCTTGTCTCGGAGCGGTATCATCAA 1246

Qy 326 nProProValGluIlyLeuLeuProLeuSerLeuIlyS 338
Db 1247 CCACCTGTAGAGAGCTTTTGGCGTGTCTTAATA 1283

RESULT 10
BT006936 1272 bp mRNA linear PRI 13-MAY-2003
LOCUS Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA,
DEFINITION complete cdb.
ACCESSION BT006936
VERSION BT006936.1 GI:30582710
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Kalinine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Cloning of human full-length cDNAs in BD Creator(TM) System Donor
vector
Unpublished
2 (bases 1 to 1272)
Kalinine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Direct Submission
Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
This cDNA clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each cDNA has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The cDNA has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
Location/Qualifiers
1..1272
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

```

/clone="GH00156X1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
1..1272
/codon_start=1
/product="carboxypeptidase B2 (plasma, carboxypeptidase
U)"
/protein_id="AAP35582.1"
/db_xref="GI:30582711"
/translation="MKLSLAVLVPIVLFCEQHVFAFQSGOVLAALPRTSRQVQLON
LTTTIELVWQPTADLIKKQVHFVNADVDNKAHLNVSIGISVLLADVDEL
QQISNDTVSPRASASYBOYSHLSNIYEFITRHPDMLTKHIGSFSEKYPYLV
LKVSGKEQAANKAIWIDCGIHAREWISPAFLWFIQTQFYGIIOYTNLLRLVDFY
VMPVNVGDYDYSWKKRNRKNSFYANNHCIGTDLNRFNFKSKHKEGASSSCSE
TYCVGLPESEPEKAVASFLRRNINOIKAVISMHSYSOHIYVPVYSYTRSKSKOHEELU
LVASEAVRALEKTSKNTRYTHGHSEFLVLAPEGGDDMIYDLGIKISFTIELRDTGY
GFLPERYIKPTCREFAAFAVSKIAHWIRNV"
BASE COUNT 373 a 268 c 273 g 358 t
ORIGIN
Alignment Scores:
Pred. No.: 8,73e-162 Length: 1272
Score: 1727.00 Matches: 337
Percent Similarity: 85.75% Conservative: 0
Best Local Similarity: 85.75% Mismatches: 1
Query Match: 96.05% Indels: 55
DB: 9 Gaps: 2
US-09-980-881A-3 (1-338) x BT006936 (1-1272)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 67 TTTCAGATGGCCAGTCTTAGTGCTCTTCTAGAACCTCTAGGCAAGTTCAGTCTTA 126
Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuIle 40
Db 127 CAGATCTTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTATT 186
Qy 41 ValIlyGlyIlyGlnValHisPheValAsnAlaSerAspValAspAsnValIlySala 60
Db 187 GTGAAGAAACCAAGTCCATTTTGTAAATGCATCTGATCGACCAATGTGAAGCC 246
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
Db 247 CATTAAATGTGAGCGGAATTCATGCAAGTGTCTGTCGCACAGCTGGGAAGATCTATT 306
Qy 81 GlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 307 CAACAGCAGATTTCACACAGACAGTCCAGCCCGAGCCTCCGATCTACTATGAACAG 366
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArgHisProAsp 120
Db 367 TATCACTCAATAAATAATCTATTCTTGATAGAAATTTATAACTGAGAGCATCTCGAT 426
Qy 121 MetLeuThrIlyIleHisIleGlySerSerPheGluIlySerPheProLeuTyrValLeuIlyS 140
Db 427 ATGCTTACAAAATCCACATTTGATCTCTCAATTTGAGAGATACCACTCTATGTTTAAAG 486
Qy 141 ValSerGlyIlyGluGlnThrAlaIlyAsnAlaIleTyrIleAspCysGlyIleHisAla 160
Db 487 GTTCTCGAAAGAACCAAGCAGCAAAAATGCCATATGGATTGACGTGGAATCCATGCC 546
Qy 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis----- 175
Db 547 AGAAGATGATCTCTCTCTTCTGCTTGTGTTTATAGGCCATATAAATCAATTCTAT 606
Qy 175 ----- 175
Db 607 GGGATAATAGGGCAATATACCAATCTCCGAGCCTGTGGATTTCTATGTTATGCCGGTG 666
Qy 176 -----AsnArgMetTyrArgIlyAsnArg 183

```

Db 667 GTTAATCGTATGTTATGACTACTCATGAGAAAGAAATCGAATGTGAGAAAGACCGT 756
Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 727 TCTTTCTATCGAACAATCATTCGATCGGACAGACCTGAATAGGAACCTTTGCTCCAAA 786
Qy 204 HisTpcCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 787 CACTGGTGTGAGGAGGTGCATCCAGTTCCTCATGCTCGGAACCTACTGTGACCTTAT 846
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 847 CCTGAGTCAGAACCCAGAGGTGAAGGAGTGCCTAGTTCCTTTGAGAGAAATATCAACCAG 906
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 907 ATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTGTTCCATATTCCTAT 966
Qy 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAlaValArg 283
Db 967 ACAGAGGTAAAGCAAGACCATGAGAACTGTCTAGTACCGAGTGAAGCAGTTCGT 1026
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1027 GCTATTGAGAAACTAGTAAATATACCATGATACATGCGCTCAGAAACCTTA 1086
Qy 304 TyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1087 TACCTAGCTCTGAGGTGGGAGCATTTGGATCTATGATTTGGCATCAATATTCGTTT 1146
Qy 324 -----ThrSerAs 326
Db 1147 ACAATTGAATCTGAGATACGGGCACATACGATTTCTGCTCGGAGCGCTTACATCAA 1206
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1207 CCCACCTGTAGAGAGCTTTTGGCGCTGTCTCAAAA 1243

RESULT 11
BT007990
LOCUS
DEFINITION Synthetic construct Homo sapiens carboxypeptidase B2 (plasma, 1272 bp mRNA linear SYN 13-MAY-2003
carboxypeptidase U) mRNA, partial cds.
ACCESSION BT007990
VERSION BT007990.1 GI:30584818
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Kallnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator (TM) System Donor vector

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1272)
AUTHORS Kallnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion (TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame.

Clone distribution: http://bioinfo.clontech.com/orfclones.
Location/Qualifiers
1..1272
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH0015611.0"
/clone_lib="BD Creator (TM) CDS Library derived from MGC collection"
/lab_host="DH5alpha T1 resistant"
/note="Vector: pDNR-Dual"
1..1272
/note="Mutations: 1271:Stop->Leu"
/codon_start=1
/transl_table=11
/product="Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U)"
/protein_id="AAP36662.1"
/db_xref="GI:30584819"
/translation="MKLCSLAVLPVLFCEOHVFAFQSGQVLAALPRTSROVQVLON LTTTVELVWQPVTDLIYKKQVHFVNASVDNVKAHLNVSGIPCSVLLADVEDLI QQQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHDMUTKIHIGSFNFKPLIYV LKVSQKEQAQNAIWDICGHAREWISPAFLWFIGHITQFYGIQYTLRLVDFY VMPVNVVDGYDYSWKKNRMNRSPFANNHCITGDLNRFASKHWCESGSSCSSE TVCGLYPESEPEVKAVASFLRNINOIKAYISMHSYOHIVFPYSYTRSKSDHELS LVASEAVRAIEKTSKNTRYTHGSGSETLVAPGGDDDIYDLGIKISFTIELRDTGT GFLLERYIKPTCREAPAFAVSKIAHVRNVL"
BASE COUNT 372 a 268 c 273 g 359 t
ORIGIN
Alignment Scores:
Pred. No.: 8,73e-162 Length: 1272
Score: 1272.00 Matches: 337
Percent Similarity: 85.75% Conservative: 0
Best Local Similarity: 85.75% Mismatches: 1
Query Match: 96.05% Indels: 55
DB: 12 Gaps: 2
US-09-980-881A-3 (1-338) x BT007990 (1-1272)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 67 TTTGAGATGGCCAGCTTCTAGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeu 40
Db 127 CAGATCTTACTACAAACATATGAGATTGTCTCTGGCAGCCGCTGACGCTTATT 186
Qy 41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
Db 187 GTGAAGAAAAACAAGTCCATTTTGTAAATGCACTGATGTCACATGTGAAAGCC 246
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 80
Db 247 CATTTAAATGTGAGCGGAATTCATGCTGTCAGTGTCTTCTGGCAGCGTGAAGATCTATT 306
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 307 CAACACAGATTTTCCAAACGACACAGTCAGCCCCGAGCCCTCCGCATCGTACTATGAACAG 366
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArgHisProAsp 120
Db 367 TATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATATACTGAGAGGCATCTGAT 426
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 427 ATGCTTACAAAAATCCACATTCGATCCTCATTTGAGAAGTACCACCTCTATGTTTTAAAG 486
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrIleAspCysGlyIleHisAla 160
Db 487 GTTCTGGAAGAAACAGCAGCAAAAATGCCATATGGATTGACCTGTGGAATCCATGCC 546
Qy 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis----- 175


```

Db      547 AGAGATGATCTCTCTGCTTCTGCTGTGCTTCTAGGCGCATATACTCAATTCTAT 606
QY      175 ----- 175
Db      607 GGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATGCCGGTG 666
QY      176 -----AsnArgMetTrpArgLysAsnArg 183
Db      667 GTTATGTGGATGTTATGACTACTCATCGAAAAAGAAATCGAATGATGGAGAAACCGT 726
QY      184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db      727 TCTTCTATGCGAACATCTTCATCGGACAGACTGAATAGGAACCTTGTCTTCAA 786
QY      204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db      787 CACTGGTGTGAGGAAGTGCATCCAGTCTCTCATGCTCGAAACCTACTGTGGACTTTAT 846
QY      224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db      847 CCTGAGTCAAGACCAAGTGAAGGAGTGGCTAGTTCTTGTGAGAGAAATATCAACAG 906
QY      244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db      907 ATTAAAGCATATCATCATGATGCTTCTTCCAGCATATAGTGTTCATATTCCTAT 966
QY      264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
Db      967 ACAGGAAGTAAAGCAAGACCATGAGGAAGTCTCTCTAGTAGCCAGTGAAGCAGTTCT 1026
QY      284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db      1027 GCTATTTCAGAAAACTAGTAAAAATACCAAGTATACAGTATACATGGCCATGGCTCAGAAA 1086
QY      304 TyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db      1087 TACCTAGTCTCTGAGGAGTGGGACGATGGATCTATGATTTGGGCATCAATATTCGTTT 1146
QY      324 -----ThrSerAs 326
Db      1147 ACATTTGAATTCGATACGGGCACATACGAGATTCTGCTCGCGAGCGTTACATCAA 1206
QY      326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db      1207 CCACCTGTAGAGAAGCTTTTGGCGCTGTCTTAAAA 1243

RESULT 12
BC007057
LOCUS
DEFINITION Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
IMAGE:3934520, mRNA, complete cds.
ACCESSION BC007057
VERSION BC007057.1 GI:13937896
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1715)
Strausberg,R.
Direct Submission
Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-x@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome

```

Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 16 Row: f Column: 21
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503004.

FEATURES

source

Location/Qualifiers

1..1715

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:12495 IMAGE:3934520"

/tissue_type="Skeletal Muscle"

/clone_lib="NIH MGC 81"

/lab_host="DH108"

/note="Vector: pDNR-LIB"

25..1296

/codon_start=1

/product="carboxypeptidase B2 (plasma)"

/protein_id="AAH07057.1"

/db_xref="GI:13937897"

/translation="MKLCLSLAVLPIVLFCQHFQFQGVLAALPRTSRQVQLQN

LTTTVEVLWQPVTDLIVKKQVHFVNASVDNVKHLNVSGIPCVLLADVEDLI

QQQINDSVTPRASASYEQVHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPLV

LKVSKEQAKNAIWDIGHAREWISFAPNCUFWIGTITQYIGIQTNLRLVDYF

VMVNVQDGYDSWKNNRMRKRNFYANHCIGTDLNNRFASKMWCSEGASSCSSE

TYCGLYPSEPEKAVASPLRRNIQIKAYISMHSYSQHVFPYSTRSKDKHELS

LVASAVRAIEKTSKNTYTHGSGETLYLARGGGDDIYDLGIKYSFTIELRDTGT

GFLPERVYIKPTCREAFSAVSKIAHVIRNV"

BASE COUNT 518 a 348 c 341 g 508 t

ORIGIN

Alignment Scores:

Pred. No.: 1,29e-161 Length: 1715

Score: 1727.00 Matches: 337

Percent Similarity: 85.75% Conservative: 0

Best Local Similarity: 85.75% Mismatches: 1

Query Match: 96.05% Indels: 55

DB: 9 Gaps: 2

US-09-980-881A-3 (1-338) x BC007057 (1-1715)

QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 Db 91 TTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 150
 QY 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuIle 40
 Db 151 CAGAATCTTACTACCAACATATGAGATTGTCTCTGGCAGCCGGTAACAGCTGACCTATT 210
 QY 41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
 Db 211 GTGAAGAAAAACAAGTCCATTTTGTGAATGCATCTGATGTCGACCAATGTGAAGCC 270
 QY 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
 Db 271 CATTTAAATGTGAGCGGAATTCATGCACTGCTTCTGCGCAGACGTGGAAGATCTTATT 330
 QY 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 Db 331 CAACAGCAGATTTCCAAACGACACAGTCAGCCCGCCGCTCCGCATCGTACTATGAACAG 390
 QY 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
 Db 391 TATCACTCACTAAATGAATCTATTCTTGATAGAAATTTATTAAGTGAAGGATCTCTGAT 450
 QY 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

QY 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis----- 175
 Db 560 AGAGAAATGGATTTCACCTGCTTCTGTTGTGGTTATAGGCTACGTGACACAAATTCAT 619
 QY 175 ----- 175
 Db 620 GGGAAAGAAATCTGTATACAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 679
 QY 176 ----- 176
 Db 680 ATGAACGTGGATGGCTATGACTACAGTGGAAAAAGAAATCGAATGTGGAGGAAGAACCGC 739
 QY 184 SerPheTyrAlaAlaSerHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
 Db 740 TCTGCTCAGAACAAACACCTCGTGGGACAGACTTCTTGAGAGAAATATCGACCAC 799
 QY 204 HistPyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
 Db 800 CACTGGTGTGAGAAAGGTGGTCAAGTTCCTCTGCTCTGAAACCTACTGTGGACTTTAT 859
 QY 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
 Db 860 CCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAGAAATATCGACCAC 919
 QY 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
 Db 920 ATTAAGCTTACATCATGATGCACTCATCTCCCAACAAATCTGTTCCCTATTCCTAT 979
 QY 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
 Db 980 ACAGAGGAAAGCAAGGACCAAGCAAGTCTCTCTAGTGGCCAGGAGGAGTTCGT 1039
 QY 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
 Db 1040 GCAATTTGAAGTATTATAAAACACACAGGTACACACAGCGAGTGGCTCAGAAAGTTTA 1099
 QY 304 TyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
 Db 1100 TATCTAGCTCTCGAGGTTCTGACGATTGGATCTATGATTGGGCATCAAAATTCGTTT 1159
 QY 324 ----- 1159
 Db 1160 ACAATTGAGCTCCGAGATACAGCAGATACGAGATTCTTGCTGCTGAGAGATACATCAA 1219
 QY 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
 Db 1220 CCCACTTGTGCAGAAAGCTTTGGCGGCATCTCTATAAA 1256

RESULT 14
 AF164524
 LOCUS
 DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)
 mRNA, complete cds.
 ACCESSION AF164524
 VERSION AF164524.1
 KEYWORDS GI:7416966
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1490)
 Marx.P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
 Characterization of mouse thrombin-activatable fibrinolysis inhibitor
 Thromb. Haemost. 83 (2), 297-303 (2000)
 JOURNAL MEDLINE
 PUBLISHED 20201996
 10739389
 REFERENCE 2 (bases 1 to 1490)
 Marx.P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
 Direct Submission
 TITLE Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100, Utrecht 3584 CX, The Netherlands

FEATURES
 source
 1. .1490
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:477035"
 1. .1490
 /gene="Tafi"
 84. .1352
 /gene="Tafi"
 /function="carboxypeptidase"
 /codon_start=1
 /product="thrombin-activatable fibrinolysis inhibitor"
 /protein_id="AAF62385.1"
 /db_xref="GI:7416967"
 /translation="MKHLGLILVAILYEQRHGFAPQSGQVLSALPRTSRQVLQNL
 TTTTVEVLWQPVTAETFEKKKEVHFVNASDVDSVKHLNVSRIIPNVLNNVEDLIE
 QOTENDTVSPRASASYEOYHSLNEIYEWIEVTECHPDLQKIYIGSSPEKYPLVL
 KYSGEQRILKNAIWDICGIIHAREWISPAELFPIGYVTFPHGKENLYRLRLHVDYFI
 MPVMVDGYDTWKQNRKNSAHKNNRCVGTDLNRFASKHWCCKGASSSCSET
 YCLXPESPEPKAVADFLRRNIDHKAYISMHSYSQQILFFPYSNRSKDHEJSL
 VASEAVRAIESINKNTRYTHGSGSESLYLAPOGSDDIYDLGIKYSFTIELRDTGRYG
 FLPLERYIKPTCAEALAAISKIVHWIRNT"
 BASE COUNT 438 a 337 c 322 g 393 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,47e-138 Length: 1490
 Score: 1494.00 Matches: 292
 Percent Similarity: 79.64% Conservative: 21
 Best Local Similarity: 74.30% Mismatches: 25
 Query Match: 83.09% Indels: 55
 DB: 10 Gaps: 2
 US-09-980-881A-3 (1-338) x AF164524 (1-1490)
 QY 1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 Db 147 TTTTCAGATGGCCAGGCTTTTATCTGCTCTTCCAGAACCTCCAGGCAAGTTCAACTACT 206
 QY 21 GlnAsnLeuThrThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuLe 40
 Db 207 CAGAAATCTTACTACACAGTATGAGTGGCTTCTCTGGCAGCCAGTGACGCTGAATCATC 266
 QY 41 ValLysLysGlnValHisPheValAsnAlaSerValAspAsnValLysAla 60
 Db 267 GAGAGAAAAGGAAGTCCACTTTTGTGAATCGCTCTGATGTCGACAGTGTCAAAGCG 326
 QY 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
 Db 327 CATTTAAATGTGAGCAGAAATTCATTTAAAGCTTCTGATGAACCAACGTTGGAGGACCTAA 386
 QY 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 Db 387 GAACAGCAGACTTTCATGACAGCGTCCGCGCTCCGCTTCATCATATGAGCAG 446
 QY 101 TyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArgHisProAsp 120
 Db 447 TATCACTCCGTAATGAATCTATCTCTGGATAGAAGTCACTAAGTGAACAGACATCTGAC 506
 QY 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
 Db 507 ATGCTCCAGAAAATCTACATCGGATCATCTCAGAAAGTACCCACTTTATGTTTTAAAG 566
 QY 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrIleAspCysGlyIleHisAla 160
 Db 567 GTCTCAGAAAGAACAAAGAAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 626
 QY 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis----- 175
 Db 627 AGAGAAATGATTTCACCTGCTTCTGTTGTGGTTATAGGCTACGTGACACAAATTCAT 686
 QY 175 ----- 175

```
Db 687 GCGAAGAAATCTGTATACAGACTTCTGAGGACGCTGATTTCTACATCATGCCCGTG 746
Qy 176 -----AsnArgMetTyrArgLysAsnArg 183
Db 747 ATGAACGCTGGATGGCTATGACTACACGTGGAAGAAATCGAATCGATGGAGGAACCGC 806
Qy 184 SerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 807 TCTGCTCACAAGAACACACCGCTCGTGGGACAGACTGGAACAGGAATTCGCTTCCAAA 866
Qy 204 HisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 867 CACTGGTGTGAGAAGGTGGCTCAAGTTCCTCTGCTGAAACCTACTGTGGACTTAT 926
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 927 CCTGAGTCTGAGCAGAGGTGAAGGAGTGGCTGACTTCTTGAGAAGAAATATCGACCAC 986
Qy 244 IleValAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 987 ATTAAGCTTACATGACTGACATCTACTCCCAAAATACTGTTTCCCTATTCCTAT 1046
Qy 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 1047 AACAGAAGCAAGCAAGGACCCAGAGAACTGTCTAGTGGCCAGGAGAGTTCGT 1106
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1107 GCAATTGAAGATTAATAAATAACACCCAGGTACACACAGGAGTGGCTCAGAAAGTTA 1166
Qy 304 TyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1167 TATCTAGCTCTGGAGGTTCTGACGATTTGATCTATGATTTGGCATCAAAATATTCGTT 1226
Qy 324 -----ThrSerAs 326
Db 1227 ACAATTGAGTCCGAGNATCAGGACGATGATTTCTGCTGCTGAGAGATACATCAA 1286
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1287 CCCACTGTGCAAGGCTTTGGCGCCATCTCTAAAA 1323

RESULT 15
AF186188
LOCUS AF186188 1430 bp mRNA linear ROD 01-OCT-1999
DEFINITION Mus musculus carboxypeptidase U mRNA, complete cds.
ACCESSION AF186188
VERSION AF186188.1 GI:6003651
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
REFERENCE 1 (bases 1 to 1430)
AUTHORS He, Y.C. and Broze, G.
TITLE Isolation and characterization of mouse liver carboxypeptidase B
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1430)
AUTHORS He, Y.C. and Broze, G.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1999) Division of Hematology, Washington
University School of Medicine, 216 S. Kingshighway, St. Louis, MO
63110, USA
FEATURES
source
1..1430
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="B6/CBAF1J"
/strain="taxon:10090"
20..1288
/note="pro-carboxypeptidase-B; thrombin-activatable
CDS
```

```
fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds
plasminogen"
/codon_start=1
/product="carboxypeptidase U"
/protein_id="AAF0528.1"
/db_xref="GI:6003652"
/translation="MKLHGLGILVAILLYEQHGFQSGQVLARSFTRTSQVOLLWL
TTTTVEVLAPQVTAETIEKKKSVHPFNASDVSVKHAHLNVSRIPNVLMNNVEDLIE
QQTFTDVSFRASASYEQYHLSNIYSEIUEVTEQHPDMLQKIYIGSFPEKYPYL
KYSGEQRIRKNAIWDICGHAHEWISPAFLWPIGVYTFQHGKENVYTRLLRHVDFYI
KPMNVYDGYDTWKKRMRNRKNSRNCVGTDLNRNFASKWCEKSGASSSCSET
YGLYPESEPEKAVADFLRNIDHTIKAVISMSHSYQQIILFPYSYNRKSKDHEELSL
VASEAVRAIGSFNKNTRYTHSGSESLEYLAPGGSDDWIYDLGIKYKSFTELRDTGRYG
FULLPERYIKPTCAEALAAISKIVHWIRNT"
BASE COUNT 421 a 320 c 310 g 379 t
ORIGIN
Alignment Scores:
Pred. No.: 1,37e-137 Length: 1430
Score: 1484.00 Matches: 290
Percent Similarity: 79.39% Conservative: 22
Best Local Similarity: 73.79% Mismatches: 26
Query Match: 82.54% Indels: 55
DB: 10 Gaps: 2
US-09-980-881A-3 (1-338) x AF186188 (1-1430)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 83 TTTTCAGATGGCCAGGCTTTTATCTGCTCTTCCAGAACTCCAGGCAAGTTCAACTACTT 142
Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuIle 40
Db 143 CAGAAATCTTACTACACAGTATGAGTGGTCTCTGGCAGCCAGTACAGCTGAATTCATC 202
Qy 41 ValLysLysGlnValHisPheValAsnAlaSerAspValAspValAsnValLysAla 60
Db 203 GAGNAGAAAAGGAAGTCCACTTTTGTGAATCGCTGCTGATGTCGACAGTGTCAAAGCG 262
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
Db 263 CATTTAAATGTGACAGAGATTCATTTTACGTTCTGATGAACAACACGTCGAGGAGCACTAAT 322
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 323 GAACAGCAGACTTTTCAATGACACGTCAGCCCCCGCTCCGCTTCTCATCTATGAGCAG 382
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArgHisProAsp 120
Db 383 TATCACTCGCTAATGAATCTATTCTGGATAGAAGTCTTAATGAACAGATCTCTGAC 442
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 443 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 502
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrIleAspCysGlyIleHisAla 160
Db 503 GTCTCAGGAAAGGAACAAAAGAAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 562
Qy 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis----- 175
Db 563 AGAAGATGGATTTTCACTGCTTTCTGTTGTGGTTTATAGGCTACGTGACACAATTCAT 622
Qy 175 ----- 175
Db 623 GGGAAAGAAAATCTGTATACAGACTTCTGAGGACGCTGGATTTCTACATCATGCCCGTG 682
Qy 176 -----AsnArgMetTyrArgLysAsnArg 183
Db 683 ATGAACGCTGGATGGCTATGACTACACGTGGAAGAAATCGAATCGATGGAGGAACCGC 742
Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
```

```
Db 743 TCTGCTCACAAGACAAACCGCTGCTGGGCACAGACCTGAACAGGAACCTTCGCTTCACAA 802
Qy 204 HistpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 803 CACTGGTGTGAGAAAGGTGGCTCAAGTTCTCTCTGCTCTGAAACCTACTGTGGACTTAT 862
Qy 224 ProGluSerGluProGluValValAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 863 CCTGAGTCTGAGCAGAGGTGAAGCGAGTGGCTGACTTCTTGAGAGAAATATCGACCAC 922
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 923 ATTAAGCTTACATCAGTATGCACTCATACTCCACAATACTGTTTCCCTATTCCCTAT 982
Qy 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 983 AACAGAAAGCAAAAGCAAGGACCAAGAAAGTGTCTCTAGTGGCCAGGAAGAGTTCGT 1042
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1043 GCAATTGGAAGTTTAAATAAAACACCAAGGTACACACACGCGAGTGGCTCAGAAAGTTA 1102
Qy 304 TyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1103 TAICTAGCTCTGGAGGTTCTGACGANTGGATCTATGATTTGGGCATCAATATTCGTTT 1162
Qy 324 -----ThrSerAs 326
Db 1163 ACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGTGCTGAGAGATACATCAAA 1222
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1223 CCCACTTGTGCAGAAAGCTTTGGCGCGCATCTCTAAAA 1259
```

Search completed: January 2, 2004, 00:16:27
Job time : 5247.6 secs

This Page Blank (uspto)

GenCore version S.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:19:56 ; Search time 57.1566 Seconds
(without alignments)
1526.015 Million cell updates/sec

Title: US-09-980-881A-3

Perfect score: 1798

Sequence: 1 FOSGQVLAALPRTSRQVQL.....IKYFSTNPVVKLLPLSLK 338

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1784	99.2	360	4 Q9P2Y6	Q9p2y6 homo sapien
2	1699.5	94.5	423	4 Q15114	Q15114 homo sapien
3	1694.5	94.2	423	4 Q961Y4	Q961y4 homo sapien
4	1487.5	82.7	422	11 Q9JHH6	Q9jhh6 mus musculus
5	1477.5	82.2	422	11 Q9QZF0	Q9qzf0 mus musculus
6	1466.5	81.6	422	11 Q9QV9	Q9eqv9 rattus norv
7	916	50.9	198	4 Q9NTI8	Q9nti8 homo sapien
8	697.5	38.8	416	13 Q9PUP2	Q9puf2 bothrops ja
9	688	38.3	408	13 Q8AXN3	Q8axn3 paralichthy
10	646.5	36.0	416	6 Q9XSP3	Q9xsp3 sus scrofa
11	627	34.9	412	11 P97597	P97597 rattus norv
12	610	33.9	437	4 Q8N4T0	Q8n4t0 homo sapien
13	606.5	33.7	434	13 Q8AVS2	Q8avs2 xenopus lae
14	587	32.6	419	13 Q8UUK1	Q8uuk1 gallus gall
15	569	31.6	374	4 Q8IVL8	Q8ivl8 homo sapien
16	558.5	31.1	436	11 Q8R4H4	Q8r4h4 mus musculus

17	528.5	29.4	419	13 Q8AXN4	Q8axn4 paralichthy
18	527.5	29.3	436	4 Q8WXQ8	Q8wxq8 homo sapien
19	521.5	29.0	419	6 Q9TV85	Q9tv85 sus scrofa
20	513.5	28.6	420	11 Q8BMK6	Q8bmk6 mus musculus
21	489	27.2	452	13 Q8UW45	Q8uw45 fugu rubrip
22	484.5	26.9	262	11 Q8AVD0	Q8avd0 mus musculus
23	477.5	26.6	247	4 Q96KZ9	Q96kz9 homo sapien
24	466.5	25.9	419	13 Q8AXN5	Q8axn5 paralichthy
25	463	25.8	1332	5 Q9W475	Q9w475 drosophila
26	461.5	25.7	422	5 Q9VLZ2	Q9vlz2 drosophila
27	449.5	25.0	226	4 Q8TDE8	Q8tdes8 homo sapien
28	444.5	24.7	351	5 Q8TQ01	Q8tq01 drosophila
29	443	24.6	584	5 Q9TZH1	Q9tzh1 caenorhabdi
30	441	24.5	351	4 Q8NA08	Q8na08 homo sapien
31	440	24.5	424	5 Q9VL87	Q9vl87 drosophila
32	440	24.5	428	5 Q8IGC4	Q8igc4 drosophila
33	426	23.7	445	5 Q9W478	Q9w478 drosophila
34	419.5	23.3	416	5 Q9VRZ3	Q9vrz3 drosophila
35	416	23.1	488	5 Q9BL88	Q9bl88 caenorhabdi
36	409	22.7	424	5 Q9VS66	Q9vs66 drosophila
37	408.5	22.7	419	5 Q9VLZ1	Q9vlz1 drosophila
38	406.5	22.6	427	5 Q9U9K2	Q9u9k2 aedes aegyp
39	406.5	22.6	430	5 Q9VL86	Q9vl86 drosophila
40	403	22.4	323	5 Q19121	Q19121 caenorhabdi
41	401	22.3	540	5 Q9TZH2	Q9tzh2 caenorhabdi
42	399.5	22.2	430	5 Q8SZP6	Q8szp6 drosophila
43	398	22.1	453	5 Q9V342	Q9v342 drosophila
44	392	21.8	423	5 Q61532	Q61532 drosophila
45	391.5	21.8	454	5 Q9TZC6	Q9tzc6 caenorhabdi

ALIGNMENTS

RESULT 1

Q9P2Y6 PRELIMINARY; PRT; 360 AA.

AC Q9P2Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B-like protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;

RT "Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.

RX MEDLINE=92042093; PubMed=1939207;

RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;

RT "Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma."

RL J. Biol. Chem. 266:21833-21838(1991).

DR HSSP; A0011969; BAA90475.1; -

DR MEROPS; M14.009; -

DR InterPro; IPR003146; Propep M14.

DR InterPro; IPR000834; Zn_carbopept.

DR Pfam; PF02244; Propep_M14; 1.

DR Pfam; PF00246; Zn_carbopept; 1.

DR PRINTS; PR00765; CRBOXYPTASE.

DR SMART; SM00631; Zn_pept; 1.

SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

Query Match 99.2%; Score 1784; DB 4; Length 360;

Best Local Similarity 99.1%; Pred. No. 6.2e-154;

Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

*his art. not
was 37aa. seq #2*

```
QY 1 FQSGQVLAALPRTSRQVQLQNLTTTVEIIVLWQVPTADLVKKQVHFFVNASDVNVKA 60
DB 23 FQSGQVLAALPRTSRQVQLQNLTTTVEIIVLWQVPTADLVKKQVHFFVNASDVNVKA 82
QY 61 HLNVSIGIPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 120
DB 83 HLNVSIGIPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 142
QY 121 MLTKIHIGSSFEKYPYLVKVSQKQAKNAIWDGIIHAREWISPAFCLWFIHITQFY 180
DB 143 MLTKIHIGSSFEKYPYLVKVSQKQAKNAIWDGIIHAREWISPAFCLWFIHITQFY 202
QY 181 KNSFYANNHCIGTDLNRNPFASVAVRAIEKTSKNTYTHGHGSETLYLAPGGDDMIYDLGIKYSF 240
DB 203 KNSFYANNHCIGTDLNRNPFASVAVRAIEKTSKNTYTHGHGSETLYLAPGGDDMIYDLGIKYSF 262
QY 241 INQIKAYISMHSYSQHIYFFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTYTHGHGS 300
DB 263 INQIKAYISMHSYSQHIYFFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTYTHGHGS 322
QY 301 ETLYLAPGGDDMIYDLGIKYSFSPNPPVEKLLPLSLK 338
DB 323 ETLYLAPGGDDMIYDLGIKYSFSPNPPVEKLLPLSLK 360
RESULT 2
Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PCPB protein.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RA "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838 (1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSP; P00730; 1PT.
DR MEROPS; M14.009; -.
DR Genew; HGNC:2300; PCPB.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboPept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboPept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
FT CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6EE79F4 CRC64;
Query Match 94.5%; Score 1699.5; DB 4; Length 423;
Best Local Similarity 89.8%; Pred. No. 3.8e-146;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 FQSGQVLAALPRTSRQVQLQNLTTTVEIIVLWQVPTADLVKKQVHFFVNASDVNVKA 60
DB 23 FQSGQVLAALPRTSRQVQLQNLTTTVEIIVLWQVPTADLVKKQVHFFVNASDVNVKA 82
QY 61 HLNVSIGIPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 120
DB 83 HLNVSIGIPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 142
QY 121 MLTKIHIGSSFEKYPYLVKVSQKQAKNAIWDGIIHAREWISPAFCLWFIHITQFY 180
DB 143 MLTKIHIGSSFEKYPYLVKVSQKQAKNAIWDGIIHAREWISPAFCLWFIHITQFY 202
QY 181 KNSFYANNHCIGTDLNRNPFASVAVRAIEKTSKNTYTHGHGSETLYLAPGGDDMIYDLGIKYSF 240
DB 203 KNSFYANNHCIGTDLNRNPFASVAVRAIEKTSKNTYTHGHGSETLYLAPGGDDMIYDLGIKYSF 262
QY 241 INQIKAYISMHSYSQHIYFFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTYTHGHGS 300
DB 263 INQIKAYISMHSYSQHIYFFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTYTHGHGS 322
QY 301 ETLYLAPGGDDMIYDLGIKYSFSPNPPVEKLLPLSLK 338
DB 323 ETLYLAPGGDDMIYDLGIKYSFSPNPPVEKLLPLSLK 360
RESULT 3
Q961Y4 PRELIMINARY; PRT; 423 AA.
AC Q961Y4 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B2 (Plasma).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007057; AAH07057.1; -.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboPept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboPept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6EE7CF5 CRC64;
Query Match 94.2%; Score 1694.5; DB 4; Length 423;
Best Local Similarity 89.5%; Pred. No. 1.1e-145;
Matches 323; Conservative 0; Mismatches 1; Indels 37; Gaps 1;
QY 1 FQSGQVLAALPRTSRQVQLQNLTTTVEIIVLWQVPTADLVKKQVHFFVNASDVNVKA 60
DB 23 FQSGQVLAALPRTSRQVQLQNLTTTVEIIVLWQVPTADLVKKQVHFFVNASDVNVKA 82
QY 61 HLNVSIGIPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 120
DB 83 HLNVSIGIPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSLNEIYSWIEFITERHPD 142
QY 121 MLTKIHIGSSFEKYPYLVKVSQKQAKNAIWDGIIHAREWISPAFCLWFIHITQFY 175
DB 143 MLTKIHIGSSFEKYPYLVKVSQKQAKNAIWDGIIHAREWISPAFCLWFIHITQFY 202
QY 176 -----NRMWRKNSFYANNHCIGTDLNRNPFASK 203
DB 203 GIIGQVNTLLRLVDFYVMPVNVVDGYDSWKKNRMWRKNSFYANNHCIGTDLNRNPFASK 262
QY 204 HWCCEGASSSSCSETCYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIYVFPYSY 263
DB 263 HWCCEGASSSSCSETCYGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIYVFPYSY 322
QY 264 TRSKSKDHEELSLVASEAVRAIEKTSKNTYTHGHGSETLYLAPGGDDMIYDLGIKYSF 323
DB 323 TRSKSKDHEELSLVASEAVRAIEKTSKNTYTHGHGSETLYLAPGGDDMIYDLGIKYSF 382
QY 324 T 324
DB 383 T 383
```

```

Db 323 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSF 382
QY 324 T 324
Db 383 T 383

RESULT 4
Q9JHH6 PRELIMINARY; PRT; 422 AA.
ID Q9JHH6
AC Q9JHH6 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE (110032P04Rik protein)
GN CPB2 OR TAFI OR I110032P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058 (2000).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=20201996; PubMed=10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor";
RL Thromb. Haemost. 83:297-303 (2000).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR EMBL; AB021968; BAB03402.1; -
DR EMBL; AF164524; AAF62385.1; -
DR EMBL; AK004045; BAB23141.1; -
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep M14.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.

```

```

SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;
Query Match 82.7%; Score 1487.5; DB 11; Length 422;
Best Local Similarity 78.1%; Pred. No. 7.4e-127;
Matches 282; Conservative 20; Mismatches 22; Indels 37; Gaps 1;

QY 1 FOSGQVLAALPRTSRQVQLNLTTTYEIVLWQPVTTADLIVKKQVHFFVNASDNDNVKA 60
Db 22 FOSGQVLSALPRTSRQVQLNLTTTYEIVLWQPVTTADLIVKKQVHFFVNASDNDNVKA 81
QY 61 HLNVSIGPCSVLLADVEDLIQOISNDTVSPRASASYEYQYHSLNIYSWIBFITERHPD 120
Db 82 HLNVSIRIPNVLNNVEDLIEQTFNDTVSPRASASYEYQYHSLNIYSWIEVITEQHPD 141
QY 121 MTKIHIGSSFEEKYPLYVLKVGSKQETAKNAIWDGCIHAREWISPAFCLWFIHG----- 175
Db 142 MLQKIYIGSSFEEKYPLYVLKVGSKQRIKNAIWDGCIHAREWISPAFCLWFIGYVTQPH 201
QY 176 -----NRMRKRSFYANNHCITGTDLNRNFASK 203
Db 202 GKENLYTRLRLRHVDYFIMPVNVVDYDTWKKNRMWRKRSFAKNRNCVGTDLNRNFASK 261
QY 204 HWCCEGASSSCSETYCGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHIIVPYSY 263
Db 262 HWCCEGASSSCSETYCGLYPESEPEVKAVADFLRNIDHIKAYISMHSYSQQLIPYSY 321
QY 264 TRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSF 323
Db 322 NRSKSDHEELSLVASEAVRAIESINKTRYTHGSGESLYLAPGGDDWIYDLGIKYSF 381
QY 324 T 324
Db 382 T 382

RESULT 5
Q9QZF0 PRELIMINARY; PRT; 422 AA.
ID Q9QZF0
AC Q9QZF0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=B6/CBAFLJ;
RC He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RT gene.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF186188; AAF00528.1; -
DR HSSP; P00730; 5CPA.
DR MEROPS; M14.009; -
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 422 AA; 48832 MW; B53FFFB09943954E CRC64;
Query Match 82.2%; Score 1477.5; DB 11; Length 422;
Best Local Similarity 77.6%; Pred. No. 6e-126;
Matches 280; Conservative 21; Mismatches 23; Indels 37; Gaps 1;

QY 1 FOSGQVLAALPRTSRQVQLNLTTTYEIVLWQPVTTADLIVKKQVHFFVNASDNDNVKA 60

```


Db 22 FQSGVLSALPRTSRQVQLQLNTTYYEVVLWQPVTAETIEKKKEVHFFVNASDVDSVKA 81
 Qy 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSHLSNIYSWIEITERHPD 120
 Db 82 HLNVSRIFFVLMNVEDLIEQQTFNDTVSPRASASYEQYHSHLSNIYSWIEITERHPD 141
 Qy 121 MLTKIHIGSSFEKYPYLVLVKVSQKEQAKNAIWDGCIHAREWISPAFCLWFIHGH 175
 Db 142 MLQKIYIGSSFEKYPYLVLVKVSQKEQAKNAIWDGCIHAREWISPAFCLWFIHGH 201
 Qy 176 -----NRMWRKNSFYANNHCIGTDLNRNPFASK 203
 Db 202 GKENLYTRLLRHVDYFIMPVMDVGYDTWKQKRMWRKNSFYANNHCIGTDLNRNPFASK 261
 Qy 204 HWCCEGASSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHVFPYSY 263
 Db 262 HWCCEGASSSCSETYCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQHVFPYSY 321
 Qy 264 TRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 323
 Db 322 NRKSKDHELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 381
 Qy 324 T 324
 Db 382 T 382

RESULT 6

ID Q9EQV9 PRELIMINARY; PRT; 422 AA.
 AC Q9EQV9;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Pre-procarboxypeptidase R.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20471387; PubMed=11021404;
 RA Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
 RA Okada N., Okada H.;
 RT "Molecular cloning and partial characterization of rat
 RT procarboxypeptidase R and carboxypeptidase N.";
 RL Microbiol. Immunol. 44:719-728 (2000).
 DR EMBL; AB042598; BAB18617.1; -.
 DR HSP; P00730; 2CTC.
 DR MEROPS; M14.009; -.
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboOpept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carboOpept; 1.
 DR SMART; SM00631; Zn_pept; 1.
 KW Carboxypeptidase.
 SQ SEQUENCE 422 AA; 48926 MW; FFFD32A5IA9366C8 CRC64;

Query Match 81.6%; Score 1466.5; DB 11; Length 422;
 Best Local Similarity 76.7%; Pred. No. 6.1e-125;
 Matches 277; Conservative 22; Mismatches 25; Indels 37; Gaps 1;

Qy 1 FQSGVLSALPRTSRQVQLQLNTTYYEVVLWQPVTAETIEKKKEVHFFVNASDVDSVKA 81
 Db 22 FQSGVLSALPRTSRQVQLQLNTTYYEVVLWQPVTAETIEKKKEVHFFVNASDVDSVKA 141
 Qy 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSHLSNIYSWIEITERHPD 120
 Db 82 YLNASRIFFVLMNVEDLIEQQTFNDTVSPRASASYEQYHSHLSNIYSWIEITERHPD 141
 Qy 121 MLTKIHIGSSFEKYPYLVLVKVSQKEQAKNAIWDGCIHAREWISPAFCLWFIHGH 175
 Db 142 MLQKIYIGSSFEKYPYLVLVKVSQKEQAKNAIWDGCIHAREWISPAFCLWFIHGH 201

Qy 176 -----NRMWRKNSFYANNHCIGTDLNRNPFASK 203
 Db 202 GKENLYTRLLRHVDYFIMPVMDVGYDTWKQKRMWRKNSFYANNHCIGTDLNRNPFASK 261
 Qy 204 HWCCEGASSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHVFPYSY 263
 Db 262 HWCCEGASSSCSETYCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQHVFPYSY 321
 Qy 264 TRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 323
 Db 322 NRKSKDHELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 381
 Qy 324 T 324
 Db 382 T 382

RESULT 7

ID Q9NTI8 PRELIMINARY; PRT; 198 AA.
 AC Q9NTI8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE BA139H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
 GN CFB2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tromans A.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL137141; CAB92622.1; -.
 DR HSP; P00730; 1CPX.
 DR MEROPS; M14.009; -.
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboOpept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carboOpept; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 KW Carboxypeptidase.
 FT NON_TER 198
 SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

Query Match 50.9%; Score 916; DB 4; Length 198;
 Best Local Similarity 99.4%; Pred. No. 2.7e-75;
 Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FQSGVLSALPRTSRQVQLQLNTTYYEVVLWQPVTAETIEKKKEVHFFVNASDVDSVKA 60
 Db 23 FQSGVLSALPRTSRQVQLQLNTTYYEVVLWQPVTAETIEKKKEVHFFVNASDVDSVKA 82
 Qy 61 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSHLSNIYSWIEITERHPD 120
 Db 83 HLNVSIGPCSVLLADVEDLIQQQISNDTVSPRASASYEQYHSHLSNIYSWIEITERHPD 142
 Qy 121 MLTKIHIGSSFEKYPYLVLVKVSQKEQAKNAIWDGCIHAREWISPAFCLWFIHGH 175
 Db 143 MLTKIHIGSSFEKYPYLVLVKVSQKEQAKNAIWDGCIHAREWISPAFCLWFIHGH 197

RESULT 8

ID Q9PUF2 PRELIMINARY; PRT; 416 AA.
 AC Q9PUF2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE Carboxypeptidase homolog.
 OS Bothrops jararaca (Jararaca).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Viperidae; Crotalinae; Bothrops.
 OX NCBI_TaxID=8724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
 RT "Screening of Bothrops jararaca pancreas cDNA library."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF190274; AAF01344.1; -
 DR HSSP; P09955; INSA.
 DR MEROPS; M14.003; -
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carboxypept; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR SMART; SM00631; Zn_pept; 1.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;
 Query Match 38.8%; Score 697.5; DB 13; Length 416;
 Best Local Similarity 38.8%; Pred. No. 5.7e-55;
 Matches 140; Conservative 66; Mismatches 112; Indels 43; Gaps 5;
 QY 1 FOSGQVLAALPRTSRQVQVQLNLTITTYEIVLWQPVTDLIVKKQVHFFVNASDVDNVKA 60
 DB 22 FDEKVVYVTRNEDEVYFLNLANIQVDFWEPDSVELVKAEMTVDFRIEADRCSEVES 81
 QY 61 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYEYQYHSLNIEYISWIEFITERHPD 120
 DB 82 ILQOGLNVEILIDNLQAVLDRLDHA---RTAGYNYEKYNSWEKIDAWTADIANENFS 138
 QY 121 MLTKIHGSSFEKYPYLVKLVKSGKEQTAKNAIWDGIIHAREWISPAFCLWFI----- 173
 DB 139 LVSRQLQITTFEGRPMPLKLV-KRPGVNVKKAIFIDCGFHAREWISPAFCQWFVREARTY 197
 QY 174 -----GHNRMKRNRSFYANNHCIGTDLNRNPFASK 203
 DB 198 GKETIMTQLLNKLDYFPLVNLIDGVYVSNKQSRMRKTRSVNAGSTCIGTDFNRNFDAA 257
 QY 204 HWCIEGASSSCSETCYGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHVFPYSY 263
 DB 258 -WCVGASRNPCSETCYGSPSEKETKALADPIRNRNIIQAYLIHYSYQMLLYPYSY 316
 QY 264 TRSKSDHELSLIVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWYDLGIKYSF 323
 DB 317 TYDLTSNNKKLNSIAKEAIREL-KVLFETTYTGGATTPAAGSGDDWAYDQGIKIAF 375
 QY 324 T 324
 DB 376 T 376
 RESULT 9
 QYAXN3 PRELIMINARY; PRT; 408 AA.
 AC QYAXN3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Carboxypeptidase B (Fragment).
 GN CPB.
 OS Paralicthys olivaceus (Flounder).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Pleuronectoidae; Paralicthidae; Paralicthys.
 OX NCBI_TaxID=8255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Srivastava A.S., Kurokawa T., Suzuki T.;

RT "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
 RT A2 and B from the Japanese flounder, Paralichthys olivaceus."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB099302; BAC53789.1; -
 FT NON TER 1
 SQ SEQUENCE 408 AA; 45904 MW; 809CAACD9D2DA34B CRC64;
 Query Match 38.3%; Score 688; DB 13; Length 408;
 Best Local Similarity 39.6%; Pred. No. 4e-54;
 Matches 143; Conservative 61; Mismatches 115; Indels 42; Gaps 5;
 QY 1 FOSGQVLAALPRTSRQVQVQLNLTITTYEIVLWQPVTDLIVKKQVHFFVNASDVDNVKA 60
 DB 12 FEGEKVRLKPVDFDEHVTILRLANSIEVDFWRPESQLVTTDIDVIRVPAIYLDKMYT 71
 QY 61 HLNVSIGPCSVLLADVEDLIQQOISNDTVSPRASAYEYQYHSLNIEYISWIEFITERHPD 120
 DB 72 LQQSDMEHVEVLIEDLQSAVDAEAD--LKPSRAHSYTKYNNLDKQVQSWIASISSNPD 128
 QY 121 MLTKIHGSSFEKYPYLVKLVKSGKEQTAKNAIWDGIIHAREWISPAFCLWFI----- 173
 DB 129 LISKQVIGNTYEGRPMTVLKLGKSSSTKPAIFMDCGIIHAREWISPAFCQWFVKEALSTY 188
 QY 174 -----GH-----NRMKRNRSFYANNHCIGTDLNRNPFASK 203
 DB 189 GSDSQWTSLLNQMDVFLVPLVFNIDGVDFTHKSNRMKTRSRKSGSCIGADPNRNF-DA 247
 QY 204 HWCIEGASSSCSETCYGLYPESEPEVKAVASFLRNINQIKAYISMHSYSQHVFPYSY 263
 DB 248 GWCMSGASNPCTDFCYTPSEIEVKVAVFIRENRSIIKAYLTVHSYQLLLFPYSY 307
 QY 264 TRSKSDHELSLIVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWYDLGIKYSF 323
 DB 308 TYQLAADHSELKMKVAEGASAL-RSLYGTTYTSGGATTPAAGSGDDWAYDGLGVKYSY 366
 QY 324 T 324
 DB 367 T 367
 RESULT 10
 QYXSP3 PRELIMINARY; PRT; 416 AA.
 AC QYXSP3;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Procarboxypeptidase B precursor (EC 3.4.17.2).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=99321926; PubMed=10391940;
 RA Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
 RA Hersberger C., Aviles F.;
 RT "Mapping the Pro-region of carboxypeptidase B by protein engineering.
 RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
 RL J. Biol. Chem. 274:19925-19933(1999).
 DR EMBL; AJ133775; CAB46991.1; -
 DR HSSP; P09955; 1PBA.
 DR InterPro; IPR003146; Propep_M14.
 DR InterPro; IPR000834; Zn_carboxypept.
 DR Pfam; PF02244; Propep_M14; 1.
 DR Pfam; PF00246; Zn_carboxypept; 1.
 DR PRINTS; PR00765; CRBOXYPTASEA.
 DR SMART; SM00631; Zn_pept; 1.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Carboxypeptidase; Hydrolase; Signal.
 FT SIGNAL 1 15 POTENTIAL.


```
QY 174 -----GHNRMWKNRSFYANNHCIGTDLNRN 199
Db 214 LLYTKSDPAMKMLNHLFYIMPVENVGVHFWNTNDRFWRKTRSRNSRFRCRGVNDRN 273
QY 200 FASHKWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSHIVF 259
Db 274 WKVK-WCEGASHPDCDDTYCGPPESEPEVKAVANFLRKHRIAYLSFHAYAQMLLY 332
QY 260 PYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGI 319
Db 333 PYSYKYATTINPRCVESAAYKAVNALQSV-YGVRYRYPASTTLYVSSGSSMDWAYKNGI 391
QY 320 KYSP 323
Db 392 PYAP 395

RESULT 13
Q8AVS2 PRELIMINARY; PRT; 434 AA.
ID Q8AVS2
AC Q8AVS2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Similar to carboxypeptidase A6.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC041315; AH41315.1; -.
KW Carboxypeptidase.
SQ SEQUENCE 434 AA; 50581 MW; B42BD77843910591 CRC64;

Query Match 33.7%; Score 606.5; DB 13; Length 434;
Best Local Similarity 36.4%; Pred. No. 11e-46;
Matches 131; Conservative 60; Mismatches 124; Indels 45; Gaps 8;

QY 5 QVLAALPRTSRQVQLNLTYYEIVLWQPVTDALIVKK--KQVHFFVNASDVNDVKAHL 62
Db 37 KVRVFPKNEREANELKDMYQQLQVDLWQSSISHIGKDTVDVHTSGNSSEV--LLTYL 94
QY 63 NVSGIPCSVLADVEDLIQ-QQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPD 120
Db 95 TKAKIQKILVNNVQSMLEAQAPRRKRKRSLSRYNYEYHPLHETESMMFTMNKTHD 154
QY 121 MLTKIHGSSFEKYPYLVKVGKEQTAKNAIWDGCIHAREWISPAFLWFI----- 173
Db 155 LVSLFTIGKSYGRSLYVLKLGKDTKSYKKAIWIDCGMHAREWIGAFQWFKAEINSY 214
QY 174 -----GH-----NRMWKNRSFYANNHCIGTDLNRNFSK 203
Db 215 NTDPMKKILNLIIYVMPVENVGVHYSWHSDFRWRKTRSKNTRYCYGVNDRNW-KV 273
QY 204 HWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSHIVFYSY 263
Db 274 HWSDEGASLPCDNTYCGPYAESEPEVKAVAQFLYKQKHVAYMSFHAYAQMLLYPSY 333
QY 264 TRSKSDHELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSF 323
Db 334 QYGAIPIFGCVESAANAVLAI-RSAYGIRYRHPASSTLYLTSGSSMDWAYNNGIPYSY 392

RESULT 14
Q8UUK1 PRELIMINARY; PRT; 419 AA.
ID Q8UUK1
AC Q8UUK1;
```

```
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Preprocarboxypeptidase A precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=domesticus; TISSUE=Pancreas;
RA Hasegawa S., Honda K., Hikami Y.;
RL "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A precursor.";
RT Nucleotide sequence of cDNA encoding chicken carboxypeptidase A
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; X64539; CAA45837.1; -.
DR MEROPS; M14.001; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxypept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_Dept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Signal; Carboxypeptidase.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 113 419 CARBOXYPEPTIDASE A.
SQ SEQUENCE 419 AA; 47006 MW; 299C334274BC9C CRC64;

Query Match 32.6%; Score 587; DB 13; Length 419;
Best Local Similarity 35.5%; Pred. No. 6.5e-45;
Matches 134; Conservative 65; Mismatches 106; Indels 72; Gaps 9;

QY 1 FQSGVLAALPRTSRQVQLNLTYYEIVLWQPVTDALIVKKQVHFFVNASDVNDV--- 57
Db 21 FVGHQVLRIVPSSDAELQVQELQE-----LEHLQDLFWLSRGLGNPVD 65
QY 58 -----VKAHLNVSGIPCSVLADVEDLI-----QQISNDTVSPRASASY-YEQYH 102
Db 66 IRVPFSLQPVKAHLBANGVPYSIMIEDVQALVDLEQMQLRRRRFVPLSTSTFDYTSYH 125
QY 103 SLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGKEQTAKNAIWDGCIHARE 162
Db 126 TLEEIVAFMDLLVAENPNLVSKLEIGRTTENRPIYVLKFS-KGKTNRPALWIDTGIHRE 184
QY 163 WISPAFLWFI-----GHNRMWKNRSFYA 187
Db 185 WYQAGSIWFAKKIVQEQDEGLANILQDMDFLEIVTNPDPGFATHTSNRMWRKTRSKRS 244
QY 188 NNHCIGTDLNRNFSKHWCEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAY 247
Db 245 GSLCVGVDPNRNW-DAGFGSGASSNPCTETHGYPANSEPEVKAIIVDFVKSHN-IKAF 302
QY 248 ISMHSYSHIVFYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGSGSETLYLAP 307
Db 303 ISHSYSQLLYPYGYTTTAVDPDKELHQVAKVAAL--SSLYGTNYKYGSIITTIYQAS 361
QY 308 GGGDDWIYDLGIKYSFT 324
Db 362 GGTIDWTYNGIKYSFT 378

RESULT 15
Q8IVL8 PRELIMINARY; PRT; 374 AA.
ID Q8IVL8
AC Q8IVL8;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Zn-carboxypeptidase.
OS Homo sapiens (Human).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RA Obaya A.J., Lopez-Otin C.;

RT "A new Zn-carboxypeptidase highly expressed in ovary.;"

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ42118; CAD19478.1; -

KW Carboxypeptidase.

SQ SEQUENCE 374 AA; 42529 MW; 404C373BB841AAD2 CRC64;

Query Match 31.6%; Score 569; DB 4; Length 374;

Best Local Similarity 39.0%; Pred. No. 2.4e-43;

Matches 110; Conservative 51; Mismatches 81; Indels 40; Gaps 5;

QY 81 QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLYL 139

DB 29 RQEIYDKSVSPWSLETYSYNIYHPMGEIYEMREISEKYEKVVTQHFLGVYIETHPMYLYL 88

QY 140 KVSQKEQTAKNAIWIDCGIHAREWISPAFLWFI----- 173

DB 89 KISQPSGNPKKIIWMDCGIHAREWIAFAFCQWFVKEILQNHKDNSSIRKLLRNLDFFVLP 148

QY 174 -----GH-----NRMRKNRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGL 222

DB 149 VLNIDGVIYTTWTDRLMRKSRSPHNGTCFGTDLNRNF-NASWCSTGASRNCODQTFCGT 207

QY 223 YPSEPEVKAVASFLRRNINQIKAYISMHSYSQHVFPYSYTRSKKHDELSLVASEAV 282

DB 208 GPVSEPETKAVASFIESKDDILCFLTMHSYGQLILTPYGYTKNKSNNHPMIQVGKAA 267

QY 283 RAIEKTSKNTRYTHGSGSETLYLAPGGDDWIYDLGIKYSFT 324

DB 268 NAL-KAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYT 308

Search completed: January 1, 2004, 19:31:56

Job time : 58.1566 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:33:22 ; Search time 369.048 Seconds
(without alignments)
2472.334 Million cell updates/sec

Title: US-09-980-881a-3
Perfect score: 1798
Sequence: 1 FOSGQVLAALPRTSRQVQL.....IKYFTSNPPVEKLLPLSLK 338

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2 1/USPTO spool/US09980881/runat 31122003 135914 26819/app query.fasta_1.1628
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0 1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -SCORES=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN 1 1 577 @runat 31122003 135914 26819 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1785	99.3	1573	21	AAC81962	Human brain carbox
2	1749.5	97.3	1625	25	ABX13670	Human protease CDN
3	1732	96.3	1272	20	AAV74302	Human plasma carbo
4	1732	96.3	1728	24	ABN95899	Gene #2397 used to
5	1732	96.3	1749	17	AAT11671	Human plasma carbo
6	1732	96.3	1749	18	AAT62846	Human plasma carbo
7	1727	96.1	1749	14	AAQ41001	Human plasma carbo
8	902	50.2	1400	21	AAQ18005	Lung cancer associ
9	646.5	36.0	1215	16	AAQ90600	Porcine Tyr-His-Me
10	637.5	35.5	1263	17	AAT42494	Human pancreatic c
11	637.5	35.5	1263	19	AAV41795	Human pancreatic c
12	637.5	35.5	1284	17	AAT42506	ProHCPB gene with
13	622	34.6	2023	25	ABZ75116	Anti-human seminal
14	619.5	34.5	1302	25	ABV77254	Nucleotide sequenc
15	619.5	34.5	1311	22	AAD03837	Human carboxypepti
16	619.5	34.5	1993	24	AAD30579	Human protease, PR
17	619.5	34.5	2154	18	AAV17331	PreproHCPB-linker
18	618	34.4	1332	24	ABU65831	Lung cancer relate
19	610	33.9	927	17	AAT35760	Rat mature carboxy
20	591	32.9	416	25	ABX49430	Bovine EST associa
21	582	32.4	921	16	AAQ90601	Porcine carboxypep
22	571.5	31.8	2128	22	AAD03839	Human carboxypepti
23	569.5	31.7	1378	25	ABT33355	NOVX DNA sequence
24	569	31.6	1125	22	AAH76477	Human secreted met
25	569	31.6	1125	22	AAH76479	cDNA encoding huma
26	569	31.6	1125	24	ABK31745	NOVX DNA sequence
27	569	31.6	1342	25	ABT33353	Human secreted met
28	569	31.6	1342	22	AAH76476	Human secreted met
29	569	31.6	1342	22	AAH76478	Human zinc-binding
30	569	31.6	1344	25	ABQ77393	Human 23566 (carbo
31	569	31.6	1603	24	AAD39053	Human protease and
32	569	31.6	1826	22	AAF81728	Mature HCPB coding
33	567	31.5	999	17	AAT42500	mature HCPB- (His)6
34	567	31.5	1053	17	AAT42497	Human carboxypepta
35	560	31.1	1053	20	AAZ24804	Modified HCPB (D25
36	560	31.1	1059	17	AAT42511	Carboxypeptidase B
37	560	31.1	1059	18	AAT62787	Modified HCPB (D25
38	559	31.1	1059	17	AAT42512	Carboxypeptidase B
39	559	31.1	1059	18	AAT62788	Carboxypeptidase B
40	556	30.9	1059	18	AAT62806	Carboxypeptidase B
41	554	30.8	1059	18	AAT62790	Carboxypeptidase B
42	554	30.8	1059	18	AAT62801	Novel human protea
43	554	30.8	1200	22	AAD05477	Carboxypeptidase B
44	553	30.8	1059	18	AAT62789	Carboxypeptidase B
45	553	30.8	1196	25	ABT33352	NOVX DNA sequence

ALIGNMENTS

RESULT 1

AAC81962

ID AAC81962 standard; cDNA; 1573 BP.

XX AAC81962;

XX 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B cDNA.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX Homo sapiens.
OS

XX	Key	Location/Qualifiers	121	MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys	140
PH	CDS	18..1100	444	ATGCTTACAAAATCCACATTGGATCTCTCATTTGAGAGTACCACCTCTATGTTTAAAG	503
FT		/*tag= a			
FT		/product= "carboxypeptidase B"			
XX	WO200066717-A1.		141	ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla	160
XX	09-NOV-2000.		504	GTTTCTGGAAAGAACACACAGCAAAATGCCATATGGATTGACTGTGGAAATCCATGCC	563
XX	01-MAY-2000; 2000WO-JP02878.		161	ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg	180
XX	30-APR-1999; 99JP-0125169.		564	AGAGAAATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGGCCATATTCGAATGTGGAGA	623
XX	(MATS/) MATSUMOTO A.		181	LysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe	200
XX	Matsumoto A;		624	AGAACCCTTCTTCTATGCGAACAATCATGTGATCGAACACACCTGAATAGCAACTTT	683
XX	WPI; 2000-687534/67.		201	AlaSerLysHisTrpCysGluGlyAlaSerSerSerCysSerGluThrTyrCys	220
XX	P-PSDB; AAB11457.		684	GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCGGAACCTACTGT	743
XX	Human brain carboxypeptidase B isolated from the hippocampus useful for screening agents for the treatment of Alzheimer's and other brain disorders		221	GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn	240
XX	Claim 2b; Page 64-68; 84pp; Japanese.		744	GGACTTTATCTCTGAGTCAGAACCCAGAGTGAAGCGAGTGGCTAGTTCTTTGAGAGAAAT	803
XX	This invention describes a novel protein with peptidase activity against brain beta-amyloid precursor protein which has been isolated from human hippocampus and which has cerebroprotective, antialzheimer's, neurotropic, neuroprotective and hemostatic activity and which can be used as a vaccine or for gene therapy. The protein, and compounds identified by screening as promoters or inhibitors of its activity, are used to regulate beta-amyloid accumulation in the brain and treat or prevent diseases in which this occurs, such as Alzheimer's, senile dementia, inherited cerebral hemorrhage, Down's syndrome, and head trauma.		241	IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro	260
XX	Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;		804	ATCAACAGATTAAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTTCCTCA	863
XX	Alignment Scores:		261	TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu	280
XX	Pred. No.:	1.31e-186	Length:	1573	
XX	Score:	1785.00	Matches:	335	
XX	Percent Similarity:	99.41%	Conservative:	1	
XX	Best Local Similarity:	99.11%	Mismatches:	2	
XX	Query Match:	99.28%	Indels:	0	
XX	DB:	21	Gaps:	0	
XX	US-09-980-881A-3 (1-338) x AAC81962 (1-1573)				
QY	1	PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu	20		
DB	84	TTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCCAAGTTCAAGTTCTA	143		
QY	21	GlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle	40		
DB	144	CAGAACTTACTACACATAGAGATTGTCTCTGGCAGCGGTAACGCTGACCTTATT	203		
QY	41	ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla	60		
DB	204	GTGAAGAAAAAACAAGTCCATTTTTTTGTAATGTCATCTGATGTCGACAAATGTGAAGCC	263		
QY	61	HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle	80		
DB	264	CATTAAATGTGACGGAATTCATGAGTGCTCTGTCGTCGACAGCTGGAAGATCTTATT	323		
QY	81	GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln	100		
DB	324	CAACAGAGATTTCACACGACACAGTCAAGCCCCGACGCTCCGATCTGCTACTATGAACAG	383		
QY	101	TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp	120		
DB	384	TATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATTAACGTAGAGGCATCTGTAT	443		
QY	Key	Location/Qualifiers	321	TyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys	338
PH	5'UTR	1..16	1044	TATTGTTTACATCAAAACCCACCTGTAGAGAAAGCTTTGGCGCTGCTCTCTATAAA	1097
FT		/*tag= a			
FT	CDS	17..1177			
FT		/*tag= b			
FT		/product= "Protease"			
FT	variation	replace(521.A)			
FT		/*tag= c			
FT		/standard name= "Single nucleotide polymorphism"			
FT	3'UTR	1178..1625			
FT		/*tag= d			

polymorphism detection; thrombotic disease; ds.

KW XX Homo sapiens.
 OS XX WO9855645-A1.
 PN XX 10-DEC-1998.
 PD XX
 XX XX 02-JUN-1998; 98WO-EP03244.
 XX XX 03-JUN-1997; 97US-0869057.
 PR XX (SCHD) SCHERING AG.
 XX XX Morser MJ, Nagashima M;
 PI XX WPI; 1999-045800/04.
 XX XX P-PSDB; AAW92270.
 DR XX
 XX XX
 PT Detecting new polymorphism of human plasma carboxypeptidase B -
 PT comprises Alanine or Threonine at position 147 of protein by DNA or
 PT protein analysis, useful to detect risk of thrombotic disease in
 PT humans
 XX XX
 PS Example 1; Page 24; 35pp; English.
 XX XX
 CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
 CC hPCPBhr147. The invention relates to a method for determining the
 CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
 CC comprises obtaining a prepared tissue or blood sample and determining the
 CC presence of DNA coding for naturally occurring polymorphs of the protein
 CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
 CC respectively). Determination of the relative distribution of the PCPB
 CC polymorphs in a patient's blood by genetic or protein analysis by the
 CC methods is useful to determine the risk of thrombotic disease in humans.
 CC Such assessments may be made by accumulating information concerning the
 CC relative distribution of the different polymorphs within the general
 CC population compared with populations known to be at risk and establishing
 CC a PCPB polymorph profile for at-risk patients.
 XX XX
 SQ Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

 Alignment Scores:
 Pred. No.: 6.72e-181 Length: 1272
 Score: 1732.00 Matches: 338
 Percent Similarity: 86.01% Conservative: 0
 Best Local Similarity: 86.01% Mismatch: 0
 Query Match: 96.33% Indels: 55
 DB: 20 Gaps: 2

 US-09-980-881A-3 (1-338) x AAV74302 (1-1272)

 QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 Db 67 TTTTCAGAGTGGCAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
 QY 21 GlnAsnLeuThrThrThrThrGluValLeuValLeuThrProValThrAlaAspLeu 40
 Db 127 CAGAACTCTTACTACCAATATGAGATTTGTTCTCTGGCAGCGGTAACAGCTGACCTTATT 186
 QY 41 ValLysLysLysGlnValHisPheValAsnAlaSerValAspAsnValLysAla 60
 Db 187 GTGAGAGAAAAACAGTCCATTTTTTTTAAATGCACTCTGATGTCGACATGTGAAAGCC 246
 QY 61 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeu 80
 Db 247 CATTTAAATGTGACGGAAATTCATGCTCTTCTGGCAGACGTGGAAGATCTTATT 306
 QY 81 GlnGlnGlnHisSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 Db 307 CAACAGAGATTTTCCAAACGACACAGTACAGCCCGCCGAGCTCCGATCGTACTATGAACAG 366
 QY 101 TyrHisSerLeuAsnGluLeuTyrSerTyrPheGluPheLeuThrGluArgHisProAsp 120

Db 367 TATCACTCACTAAATGAAATCTATTCTTGATAGAAATTTATACTGAGAGCATCTCTGAT 426
 QY 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
 Db 427 ATGCTTTACAAAAATCCACATTTGATCTCTCATTTGAGAAGTACCACCTCTATGTTTTAAAG 486
 QY 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrPheIleAspCysGlyIleHisAla 160
 Db 487 GTTCTCTGGAAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
 QY 161 ArgGluTyrPheSerProAlaPheCysLeuTyrPheIleGlyHis----- 175
 Db 547 AGAGATGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAT 606
 QY 175 ----- 175
 Db 607 GGGATAATAGGGCAATATATACCAATCTCTCAGGCTTGTGATTTCTATGTTATGCCGGTG 666
 QY 176 -----AsnArgMetTyrArgLysAsnArg 183
 Db 667 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAAATCGAATGTGGAGAAAGACCGT 726
 QY 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
 Db 727 TCTTTCTATGGAAACAATCATTTGCATCGGACAGACCTGAATAGGAACCTTTGCTTCCAA 786
 QY 204 HisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
 Db 787 CACTGGTGTGAGGAGGTGCATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGACTTTAT 846
 QY 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
 Db 847 CCTGAGTCAGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 906
 QY 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
 Db 907 ATTAAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTTCATATTCCTAT 966
 QY 264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
 Db 967 ACACGAAGTAAAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCATTCGT 1026
 QY 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
 Db 1027 GCTATTGAGAAAACTAGTAAAAATACAGGTATACATGCGCCATGGCTCAGAAACCTTA 1086
 QY 304 TyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
 Db 1087 TACCTAGCTCTCGAGGTGGGACGATTGGATCTATGATTTGGGCATCAATATTCGTTT 1146
 QY 324 -----ThrSerAs 326
 Db 1147 ACAATTGAATCTCGAGATACGGGCACATACGGAATCTTGTCTCGGAGCGGTTACATCAA 1206
 QY 326 nProProValGluLysLeuProLeuSerLys 338
 Db 1207 CCACCTCTAGAGAAGCTTTTGGCGCTGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1243

 RESULT 4
 ABN95899
 ID ABN95899 standard; DNA; 1728 BP.
 XX
 AC ABN95899;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Gene #2397 used to diagnose liver cancer.
 XX
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX

OS Homo sapiens.
 PN W0200229103-A2.
 XX
 XX
 PD
 XX
 XX
 PF 02-OCT-2001; 2001WO-US30589.
 XX
 XX 02-OCT-2000; 2000US-237054P.
 FR
 XX (GENE-) GENE LOGIC INC.
 PA
 XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 PI
 XX WPI; 2002-426119/45.
 DR
 XX
 XX Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample -
 XX
 XX Claim 1; SEQ ID NO 2397; 298pp; English.
 PS
 XX The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumour in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytosstatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.06e-180 Length: 1728
 Score: 1732.00 Matches: 338
 Percent Similarity: 86.01% Conservative: 0
 Best Local Similarity: 86.01% Mismatches: 0
 Query Match: 96.33% Indels: 55
 DB: 24 Gaps: 2
 US-09-980-881A-3 (1-338) x ABN95899 (1-1728)
 QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 DB 86 TTTTCAGAGTGGCCAAAGTTCCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145
 QY 21 GlnAsnLeuThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 40
 DB 146 CAGAATCTTACTCAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATT 205
 QY 41 ValIysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
 DB 206 GTGAAGAAAAAACAAGTCCATTTTTTTGTAATGTCATCTGATGTCGACCAATGTGAAGCC 265
 QY 61 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
 DB 266 CATTTAATGTAGCGGAATTCATGTCAGTGTCTGTCGACAGCTGGGAAGATCTTATT 325
 QY 81 GlnGlnGlnLeuSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 DB 326 CAACAGCAGATTTTCAACAGCACAGTACAGCCCCGAGCTCCGATCGTACTATGAACAG 385
 QY 101 TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheLeuThrGluArgHisProAsp 120
 DB 386 TATCACTCACTAAATGAATCTATTCTTGGATAGAATTTATACTAGAGGCATCCTGAT 445
 QY 121 MetLeuThrIysIleHisIleGlySerSerPheGlyLysTyrProLeuTyrValLeuLys 140
 DB 446 ATGCTTTACAAAAATCCACATTTGGATCTCTATTGGAGAAGTACCACCTCTATGTTTTAAAG 505
 QY 141 ValSerGlyLysGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
 DB 506 GTTTCCTGGAAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAAATCCATGCC 565
 QY 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
 DB 566 AGAGAAATGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 625
 QY 175 ----- 175
 DB 626 GGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 685
 QY 176 -----AsnArgMetTrpArgLysAsnArg 183
 DB 686 GTTAAATGTGGACGGTTATGACTACTCATCGAAAAAGAAATCGAATGTGGAGAAAGAACCGT 745
 QY 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
 DB 746 TCTTTCTATGCGAAACAATCATTCATCGGAACAGACCTGAATAGGAACCTTTGCTTCCAAA 805
 QY 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
 DB 806 CACTGGTGTGAGGAAGTGCATCCAGTCTCTCATGCTCGGAAACCTACTCTGGACTTTAT 865
 QY 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
 DB 866 CCTGAGTCAGAACAGAGTGAAGGAGTGGCTAGTCTTCTTGAGAGAAATATCAACACAG 925
 QY 244 IleIysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
 DB 926 ATTAAGCATACATCAGCAGTCACTTCTATCTCCAGCATATAGTGTTCCTATATTCCTAT 985
 QY 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
 DB 986 ACAGGAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTACCCAGTGAAGCAGTTCTGT 1045
 QY 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
 DB 1046 GCTATTGAGAAAACTAGTAAAAATACCAAGTATACACATGCGCCATGGCTCAGAAACCTTA 1105
 QY 304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
 DB 1106 TACCTAGCTCTCTGGAGGTGGGACGATTGGATCTATGATTGGGCATCAAAATATTCGTTT 1165
 QY 324 -----ThrSerAs 326
 DB 1166 ACAATTGNACTTCGAGNATCGGGCACATACGGATTCTTGCTCGCGGAGCGGTATACATA 1225
 QY 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
 DB 1226 CCCACCTGTAGAAGCTTTTGGCGCTGTCTCTAAAA 1262
 RESULT 5
 AAT11671
 ID AAT11671 standard; DNA; 1749 BP.
 XX
 AC AAT11671;
 XX
 XX 25-MAR-2003 (updated)
 DT 12-APR-1996 (first entry)
 XX
 XX Human plasma carboxypeptidase B coding sequence.
 DE
 XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
 KW purification; plasminogen; affinity column; ss.
 XX
 OS Homo sapiens.

```

XX Key Location/Qualifiers
FH CDS 41..1312
FT /*tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT /*tag= b
FT mat_peptide 107..1309
FT /*tag= c
XX
PN US5474901-A.
XX
XX 12-DEC-1995.
XX
XX 19-JUL-1994; 94US-0277540.
XX
XX 01-FEB-1991; 91US-0649591.
XX
XX 14-OCT-1992; 92US-0959944.
XX
XX 15-DEC-1993; 93US-0167727.
XX
XX 19-JUL-1994; 94US-0277540.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1996-039508/04.
XX
XX P-PSDB; AAR90293.
XX
XX Antibody to human plasma carboxypeptidase B - useful for detecting
XX and purifying hPCPB for use in treating clotting disorders e.g.
XX haemophilia A
XX
XX Disclosure; Figure 4; 40pp; English.
XX
XX An antibody which specifically binds human plasma carboxypeptidase B
XX (hPCPB) and does not cross react with other carboxypeptidases is
XX useful for the detection of hPCPB in vitro. The antibody is also
XX used for purifying hPCPB from a sample. Purification comprises
XX passing a sample thought to contain hPCPB over either a column to
XX which antibody has been bound, or a plasminogen affinity column,
XX eluting the column and then recovering the fraction containing the
XX hPCPB.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1.08e-180 Length: 1749
XX Score: 1732.00 Matches: 338
XX Percent Similarity: 86.01% Conservative: 0
XX Best Local Similarity: 86.01% Mismatches: 0
XX Query Match: 96.33% Indels: 55
XX DB: 17 Gaps: 2
XX
XX US-09-980-881A-3 (1-338) x AAT11671 (1-1749)
XX
XX 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
XX 107 TTTTCAGAGTGCCCAAGTCTAGTGTCTTCTAGAACCTCTAGGCAAGTCAAGTTCTA 166
XX
XX 21 GlnAsnLeuThrThrThrTyrrGluileValLeuTrrGlnProValThrAlaAspLeuile 40
XX 167 CAGNACTTCTACTCAACATAGATGATTTGTTCTTGGCAGCGGTAAACGCTGACCTATT 226
XX
XX 41 ValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
XX 227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAATGCACTGATGTCGACAAATGTGAAGCC 286
XX
XX 61 HisLeuAsnValSerGlylleProCysSerValLeuLeuAlaAspValGluAspLeuile 80
XX 287 CATTTAAATGTGACCGAATTCATGTCAGTGTCTTGTGTCGACAGCGTGAAGATCTTTATT 346
XX
XX 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrrTyrrGluGln 100

```

```

Db 347 CAAACAGCAGATTTCACACGACAGTCAGCCCGCCGAGCTCCGCTACTATGAACAG 406
Qy 101 TyrHisSerIeuAsnGluileTyrrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 407 TATCACTCACTAAATGAATCTATTCTTGATAGAAATTTATTAACAGAGGATCTCTGAT 466
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrrProLeuTyrrValLeuLys 140
Db 467 ATGCTTACAAAAATCCACATTTGATCTCATTTGAGAGTACCCACTCTATGTTTAAAG 526
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlylleHisAla 160
Db 527 GTTTCGTGAAAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
Db 587 AGAGAAATGGATCTCTCCTCTTCTGCTTGTGTTTCATAGGCCATATAAACAATCTAT 646
Qy 175 ----- 175
Db 647 GGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATGCGGTG 706
Qy 176 -----AsnArgMetTrpArgLysAsnArg 183
Db 707 GTTAAATGTGACGGTTATGACTACTCATCGAAAAAGAAATCGAATGTGGAGAAAGACCGT 766
Qy 184 SerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 767 TCTTTCTATCGGAACAATCATTTGCTCGGAACAGACCTGAATAGGAACCTTTGCTCCAA 826
Qy 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrrCysGlyLeuTyrr 223
Db 827 CACTGTGTGTGAGGAGGTGATCCAGTTCTCATGCTCGAAACCTTCTGTGGACTTTAT 886
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 887 CCTGAGTCAGAACACAGAGTGAAGCGAGTGGCTAGTTCTTGAGAGAAATAATCAACCAG 946
Qy 244 IleLysAlaTyrrIleSerMetHisSerTyrrSerGlnHisIleValPheProTyrrSerTyrr 263
Db 947 ATTAAGACATACATCAGCATGCTATCCATCTCCAGCATATAGTGTTCATATTCCTAT 1006
Qy 264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 1007 ACACGAAGTAAAGCAAGAACACCATGAGGAACGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrrThrHisGlyHisGlySerGluThrLeu 303
Db 1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126
Qy 304 TyrLeuAlaProGlyGlyAspAspTrpIleTyrrAspLeuGlylleLysTyrrSerPhe 323
Db 1127 TACCTAGCTCCTGGAGGTGGGACGATTCGATCTATGATTTGGGCATCAAAATTCGTTT 1186
Qy 324 -----ThrSerAs 326
Db 1187 ACAATTGAACCTCGAGATAGGGCACATAGGATTCCTGCTCCGAGCGTTACATCAA 1246
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1247 CCCACCTGTAGAGAAGCTTTTGGCGCTGCTCTCTAAAA 1283
XX
XX RESULT 6
XX AAT62846
XX ID AAT62846 standard; DNA; 1749 BP.
XX AC AAT62846;
XX XX
XX 25-MAR-2003 (updated)
XX DT 08-MAY-1997 (first entry)
XX XX
XX Human plasma carboxypeptidase B coding sequence.
XX DE

```

XX Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
 KW plasma; plasminogen; ss.
 XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 41..1312
 FT /*tag= a
 FT /product= Human PCPB
 FT sig_peptide 41..106
 FT /*tag= b
 FT mat_peptide 107..1309
 FT /*tag= c
 FT misc_binding 134..177
 FT /*tag= d
 FT /bound_moeity= 46_bp_probe

FT US5593674-A.

PD 14-JAN-1997.

XX 27-APR-1995; 95US-0430787.

XX 01-FEB-1991; 91US-0649591.

PR 14-OCT-1992; 92US-0959944.

PR 15-DEC-1993; 93US-0167727.

PR 19-JUL-1994; 94US-0277540.

PR 27-APR-1995; 95US-0430787.

XX (GETH) GENENTECH INC.

XX Dryna DT, Eaton DL;

XX WPI; 1997-099413/09.

DR P-PSDB; AAW14733.

XX Using human plasma carboxypeptidase B in blood coagulation - is
 PT functionally related to carboxypeptidase A and pancreas
 PT carboxypeptidase B

XX Example 2; Column 37-42; 39pp; English.

XX This sequence encodes human plasma carboxypeptidase B (PCPB) which
 CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
 CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
 CC purified from human plasma or by transformed cell culture by
 CC extraction using plasminogen bound to a solid phase.
 CC (Updated on 25-MAR-2003 to correct PF field.)

XX SQ Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

XX Alignment Scores:

Pred. No.: 1.08e-180 Length: 1749
 Score: 1732.00 Matches: 338
 Percent Similarity: 86.01% Conservative: 0
 Best Local Similarity: 86.01% Mismatches: 55
 Query Match: 96.33% Indels: 2
 DB: 18 Gaps: 2

US-09-980-881A-3 (1-338) x AAT62846 (1-1749)

QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 DB 107 TTTGAGTGGCCAGTTCTAGTGTCTTCTTAGAACCTCTAGGACAGTTCAAGTTCTA 166
 QY 21 GlnAsnLeuThrThrThrThrGluileValLeuTrpGlnProValThrAlaAspLeuile 40
 DB 167 CAGAATCTTACTACATATGAGATTGTTCTCTGGCAGCGCGTAACAGCTGACCTTATT 226
 QY 41 VallylsylsGlnValHisPhePheValAsnAlaSerAspValAsnValysAla 60
 DB 227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACATGTGAAAGCC 286

RESULT 7

AAQ41001

ID AAQ41001 standard; cDNA; 1749 BP.

XX

QY 61 HisLeuAsnValSerGlyileProCysSerValLeuLeuAlaAspValGluAspLeuile 80
 DB 287 CATTTAAATGTGAGCGGAATTCATGTCAGTGTCTTCTGTCGACAGTGAAGATCTTATT 346
 QY 81 GlnGlnGlnileSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 DB 347 CAACAGCAGATTTCNAACGACACAGTCAGCCCGGAGCCTCCGCATGCTACTATGAAACAG 406
 QY 101 TyrHisSerLeuAsnGluileTyrSerTrpIleGluPheileThrGluArgHisProAsp 120
 DB 407 TATCACTCAATAATGAAATCTATTCTTGATAGAAATTTATTAACAGAGGCATCTGAT 466
 QY 121 MetLeuThrLysIleHisileGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
 DB 467 ATGCTTTACAAAAATCCACATTCGATCTCTCATTTGAGAAGTACCACCTCTATGTTTTAAAG 526
 QY 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
 DB 527 GTTCTGGAAAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGAAATCCATGCC 586
 QY 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheileGlyHis----- 175
 DB 587 AGAGAATGATCTCTCTGCTTCTGCTTGTGCTTCTCATAGGCCATATACTCAATTCTAT 646
 QY 175 ----- 175
 DB 647 GGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTCTATGTTATGCCGGTG 706
 QY 176 -----AsnArgMetTrpArgLysAsnArg 183
 DB 707 GTTAATGTGGCGGTTATGACTACTCATGGAAGAAAGAAATCGAATGTGGAGAAAGACCGT 766
 QY 184 SerPheTyrAlaAsnAsnHisCysileGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
 DB 767 TCTTTCTATGCGNACAACTCATTCGATCGGACAGACTGAAATAGGAACCTTGTCTCCAAA 826
 QY 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
 DB 827 CACTGGTGTGAGGAAGTGCATCCAGTCTCTCATGCTCGGAAACCTACTGTGGACTTAT 886
 QY 224 ProGluSerGluProGluVallylsAlaValAlaSerPheileuArgAsnileAsnGln 243
 DB 887 CTGAGTCAGAACCCAGAGAGTGAAGGAGTGGCTAGTCTTCTTGAGAGAAATATCAACCAG 946
 QY 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
 DB 947 ATTAAAGCATACATCAGCATGCTTCTACTCCAGCATATAGTGTTCATATTTCTAT 1006
 QY 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
 DB 1007 ACACGAAGTAAAGCAAGAACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCAGTTCGT 1066
 QY 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
 DB 1067 GCTATTGAGAAACTAGTAAAAATACCAAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126
 QY 304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
 DB 1127 TACCTAGTCTCTGGAGGTGGGACGATTGGAATCATGATTTGGGCATCAANTATTCGTTT 1186
 QY 324 -----ThrSerAs 326
 DB 1187 ACAATTGAATCTCGAGATACGGGCACATACGGATTCTTGTGCGGAGCGCTTATACAAA 1246
 QY 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
 DB 1247 CCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 1283

AAQ41001;
 24-AUG-1993 (first entry)
 Human plasma carboxypeptidase B gene.
 PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
 Homo sapiens.
 Key Location/Qualifiers
 sig_peptide 41..106 /*tag= a
 mat_peptide 107..1312 /*tag= b
 /*product= PCPB
 misc_feature 134..177 /*tag= c
 /*function= probe
 /*note= "used to obtain full-length clones"
 US5206161-A.
 27-APR-1993.
 01-FEB-1991; 91US-0649591.
 01-FEB-1991; 91US-0649591.
 (GETH) GENENTECH INC.
 Drayna DT, Eaton DL;
 WPI; 1993-151724/18.
 P-PSDB; AAR36273.
 New human plasma carboxypeptidase B - used as haemostatic regulator for clotting blood, partic. for treating blood clotting disorders, e.g. haemophilia
 Disclosure; Fig 4; 40pp; English.
 Human plasma carboxypeptidase B was isolated from human plasma and partially sequenced. Oligonucleotide primers were designed based on the partial amino acid sequences. The primers were used in a PCR amplification to identify cDNA encoding PCPB from a human liver cDNA library. The PCR product was capable of encoding the first 37 amino acids of PCPB; a 46mer probe was used to obtain the full-length sequence which, although disclosed in the specification, is not claimed. PCPB inhibits the enzymatic conversion by tPA of plasminogen to plasmin in the presence of fibrinogen.
 Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;
 Alignment Scores:
 Pred. No.: 3 85e-180 Length: 1749
 Score: 1727.00 Matches: 337
 Percent Similarity: 85.75% Conservative: 0
 Best Local Similarity: 85.75% Mismatches: 1
 Query Match: 96.05% Indels: 55
 DB: 14 Gaps: 2
 US-09-980-881A-3 (1-338) x AAQ41001 (1-1749)
 Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 107 TTTCAGAGTGGCCAAAGTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166
 Qy 21 GlnAsnLeuThrThrThrGluLeuValLeuTrpGlnProValThrAlaAspLeu 40
 167 CAGAACTCTTACTACAAACATATAGATTGTTCTCTGGCAGCGGGTAACAGCTGACCTTTATT 226
 Qy 41 VallysLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60

227 GTGACAGAAAAACAAGTCCATTTTTTTGTAATGTCATCTGATGTCACAATGTGAAAGCC 286
 Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 80
 287 CATTTAAATGTGAGCGGAATTCATGACGTGCTTCTGCTGGCAGACGTGAAGATCTTATT 346
 Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 347 CAACAGCAGATTTCACACGACACAGTCAGCCCCCGAGCCCTCCGCATGCTACTATGACAG 406
 Qy 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
 407 TATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATAACTGAGAGGCATCTGAT 466
 Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
 467 ATGCTTTACAAAATCCACATTTGGATCTCTCATTTGAGAAGTACCACCTCTATGTTTTAAAG 526
 Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
 527 GTTTCGGAAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGATCCATGCC 586
 Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
 587 AGAGATGGATCTCTCTGCTTCTGCTTGTGTTTCATAGGCCATATAACTCAATTCAT 646
 Qy 175 ----- 175
 647 GGGATAATAGGGCAATATACCAATCTCTCAGGCTTGTGGATTCTTATGTTATGCCGGTG 706
 Qy 176 -----AsnArgMetTrpArgLysAsnArg 183
 707 GTTAATGTGACGGTTATGACTACTCATCGAAAAGAAATCGAATGTGGAGAAGAACCGT 766
 Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
 767 TCTTTCTATGCGAACAAATCATTCATCGGAACAGACCTGAATAGGAACCTTTGCTTCAAA 826
 Qy 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
 827 CACTGTGTGAGGAGGTGCATGTCAGTCTCTCATGCTCGAAACCTACTGTGGACTTTAT 886
 Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
 887 CTTGAGTCAAGAACCAAGTGAAGGAGGTGGCTAGTTCTTGAGAAGAAATATCAACCAG 946
 Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
 947 ATTAAAGCATACATCAGATGCAATTCATCTCCAGCATATAGTGTTCATATTCCTAT 1006
 Qy 264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
 1007 ACACGAAGTAAAGCAAGAACCATGAGGAACCTGCTCTAGTAGCCAGTGAAGCAGTTCGT 1066
 Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
 1067 GCTATTGAGAAAACCTAGTAAAAATACAGGTATACAGTGGCCATGGCTCAGAAACCTTA 1126
 Qy 304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
 1127 TACCTAGCTCTGGAGGTGGGACGATGGATCTATGATTGGGCATCAATATTCGTTT 1186
 Qy 324 -----ThrSerAs 326
 1187 ACAATTGAATCTCGAGATACGGGCACATACGGATTCCTGCTCGGAGCGCTTACATCAA 1246
 Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
 1247 CCACCTCTAGAGAAGCTTTTCCGCTGTCTCTTCTTAAAA 1283
 RESULT 8
 AAF18005

ID AAF18005 standard; DNA; 1400 BP.
 AC AAF18005;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polynucleotide sequence SEQ ID 24.
 XX
 KW Human; lung cancer associated protein; neuroprotective; cytotostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotropic; antinfetive; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 XX WPI: 2000-587514/55.
 DR P-PSDB; AAB58129.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 507; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytotostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotropic; antinfetive; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer.
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;
 Alignment Scores:
 Pred. No.: 2.88e-89 Length: 1400
 Score: 902.00 Matches: 181
 Percent Similarity: 76.05% Conservative: 0
 Best Local Similarity: 76.05% Mismatches: 1
 Query Match: 50.17% Indels: 56
 DB: 21 Gaps: 2
 US-09-980-881A-3 (1-338) x AAF18005 (1-1400)
 QY 157 GlyIleHisAlaArgGluThrPheSerProAlaPheCysLeuTrpPheIleGlyHis--- 175
 Db 10 GGAATCCATCCAGANAATGATCTCTCTGCTTTCTGCTGTGGTTTCATAGGCCATATA 69

QY 175 ----- 175
 Db 70 ACTCAATTTATGGGATAATAGGGCAATATACCAATCTCTAGGCTTGTGGATTTCTAT 129
 QY 176 -----AenArgMetTr 179
 Db 130 GTTATGCCGGTGGTTAATGTGGATGGTTATGAACACTACTCATGGAAAAAGAAATGATGTG 189
 QY 179 pArgLysAenArgSerPheTyrAlaAenAenHisCysIleGlyThrAspLeuAenArgAs 199
 Db 190 GAGAAAGAACCGTCTCTTCTATCGAACAATCATTCATCGGAACAGACCTGTAATAGGAA 249
 QY 199 nPheAlaSerLysHisTyrCysGluGlyAlaSerSerSerSerCysSerGluThrTy 219
 Db 250 CTTTGTCTTCAACACTGCTGTGAGGAAGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309
 QY 219 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaLaserPheLeuArgAr 239
 Db 310 CTGTGGACTTTTATCTCTGAGTCAGAACACAGAGTGAAGGAGTGGTAGTTTCTTGAGAAG 369
 QY 239 gAenIleAenGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 259
 Db 370 AATATCAACACAGATTAAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGT 429
 QY 259 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 279
 Db 430 TCCATATTCCTATACACAGAAAGTAAAGACCAAGACCATGAGGAACACTGTCTCTAGTAGCCAG 489
 QY 279 rGluAlaValArgAlaIleGluLysThrSerLysAenThrArgTyrThrHisGlyHisG1 299
 Db 490 TGAAGCAGTTCGTGCTATTGAGAAACTAGTAAATAATACCAGGTATACACATGGCCATGG 549
 QY 299 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrrPileTyrAspLeuGlyI1 319
 Db 550 CTCAGAAACCTTATACCTAGCTCTCGAGGTGGGGACGATTGGATCTATGATTTGGGCAT 609
 QY 319 elysTyrSerPhe----- 323
 Db 610 CAAATATTCGTTTACCAATTTGAACCTTCGAGATACGGGCACATACGGATTCTTGTGCGCGA 669
 QY 324 -----ThrSerAenProValGluLysAlaLeuProLeuSerLeuLys 338
 Db 670 GCCTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAAAA 719
 RESULT 9
 AAF18005
 ID AAF18005 standard; DNA; 1215 BP.
 AC AAF18005;
 DT 13-MAR-1996 (first entry)
 DE Porcine Tyr-His-Met procaryoxypeptidase B coding sequence.
 XX Porcine Tyr-His-Met procaryoxypeptidase B coding sequence.
 KW Procaryoxypeptidase B; carboxypeptidase B; Pichia; pCPB;
 KW human serum albumin; premating factor alpha; mating factor alpha;
 KW proCB; ds.
 OS Sus scrofa.
 FH Key Location/Qualifiers
 CDS 1..1215
 FT /tag= a
 FT /product= Porcine procaryoxypeptidase B.
 XX
 PN WO9514096-A1.
 XX
 PD 26-MAY-1995.
 XX
 PF 16-NOV-1994; 94WO-US13142.
 XX
 PF 16-NOV-1993; 93US-0153258.
 XX

```

PA (BLIL ) LILLY & CO ELI.
XX
PI Fayerman JT, Greenen DP, Hersherberger CL, Larson JL;
PI Sterner JL, Zhang H;
XX
DR WPI; 1995-200386/26.
DR P-PSDB; AAR75131.
XX
PT DNA encoding porcine carboxypeptidase B - used for transforming
PT host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS Claim 2; Page 20-21; 34pp; English.
XX
CC The porcine carboxypeptidase B coding sequence can be place in a
CC bacterial or pref. Pichia yeast expression vector. The expression
CC vector further comprises the signal peptide of either human serum
CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC pFJ474 - NRRL B-21032) or the porcine proCBP signal peptide,
CC (designated pLGD27 - NRRL B-21027). The method can be used for
CC producing large amounts of porcine carboxypeptidase B and when
CC produced in Pichia yeast, the protein does not need solubilisation
CC or folding. The produced enzyme is then used for pref. cleaving
CC basic residues from the carboxy terminus of proteins.
XX
SQ Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;

Alignment Scores:
Pred. No.: 3,6e-61 Length: 1215
Score: 646.50 Matches: 137
Percent Similarity: 54.57% Conservative: 60
Best Local Similarity: 37.95% Mismatches: 121
Query Match: 35.96% Indels: 43
DB: 16 Gaps: 6

US-09-980-881A-3 (1-338) x AA090600 (1-1215)
QY 1 PheGlnSerGlyInValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 28 TTCGAGGGGAGGAGGTTCCTCGTCAATGTTGAAGATGAAATGACATCAGCTTACT 87
QY 21 GlnAsnLeuThrThrThrGluLeuValLeuProValThrAlaAspLeuLe 40
Db 88 CATGAGTTGGCCAGCACCAGCAGATTGACTTCTGGAAACACAGATTCTGTACACAATC 147
QY 41 ValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60
Db 148 AAACCTCACAGTACAGTTGACTTCCGTGCGTGAAGCAGAGATATTTGGCTGTGGAAGAC 207
QY 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
Db 208 TTTCTGGAGCAGATGAACATACAAATATGAGGTACTATAAACAACCTGAGATCTGTGCTC 267
QY 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrtyrGluGln 100
Db 268 GAGGCTCAGTTTGACACGACAGTC-----CGTACAACCTGGACACAGATTATGAGAAG 318
QY 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 319 TACAACAACCTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAATAATCCAGAC 378
QY 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 379 CTCATCTCTCGCAGCCATCGGAACCTACATATTTAGGAAACAATATATACCTCCCTCAAG 438
QY 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 439 GTT---GGCAAACTGGACCAATAAGCCTGCCATTTTCATGAGCTGTGTTCCATGCC 495
QY 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
Db 496 AGAGATGGATTTCCCATGATTTTGGCAGTGGTTTGTGAGAGAGGGTGTCTCACCTAT 555

```

173 ----- 173

556 GGATATGAGATGACATGACAGATTCCTCAACAGCTAGACTTTTATGCTTTCCTGTG 615

174 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 183

616 CTCAATATTGATGGCTTACATCTACACCTGGACCCAGAACCGAATGTGGAGAAAGACCGC 675

184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203

676 TCTACCAATGCTTGAATCTACCTGCAATGGCAGACCCCAACAGAAATTTT---GATGCT 732

204 HisTrpCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGlyLeuTyr 223

733 GGTGTGTGACAACTGGAGCTCTCAGACCCCTGGATGAGACTTACTGTGGATCTGCT 792

224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIleAsnGln 243

793 GCAGAGTCTGMAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACACCTCTCTCC 852

244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263

853 ATCAAAAGCATACCTGACGATCCACTCATCTACAGATGATACCTACCTTATTCCTAT 912

264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283

913 GATTACAACCTCCCGAGAACAAATGCTGATGATTAACCTGGCTAAGGCTGCGTGAAA 972

284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303

973 GAACCTT---GCTACACTGTATGCGACCAAGTACACATACGCGCCAGGAGCTACAACAATC 1029

304 TyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323

1030 TATCTGCTGCTGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTTC 1089

324 Thr 324

1090 ACC 1092

RESULT 10

AAAT42494

ID AAT42494 standard; DNA; 1263 BP.

XX AC AAT42494;

XX 12-FEB-1997 (first entry)

DE Human pancreatic carboxypeptidase B encoding sequence.

XX ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;

KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;

KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;

KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;

KW Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..1248

FT /tag= a

FT /product= preproenzyme

FT sig_peptide 1..39

FT /tag= b

FT /note= "partial enzyme secretion leader sequence (pre-sequence); residues -108 to -96"

FT sig_peptide 40..324

FT /tag= c

FT /note= "pro-sequence; residues -95 to -1"

FT mat_peptide 385..1245

FT /tag= d

FT /note= "mature enzyme; encodes residues +1 to +307"


```
PN WO9620011-A1.
XX
PD 04-JUL-1996.
XX
PF 21-DEC-1995; 95WO-GB02991.
XX
PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
PA (ZONE ) ZENBCA LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI Hennam JF, Hennequin LFA, Mareham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WP1; 1996-321650/32.
DR P-PSDB; AAW06172.
XX
PT Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
PT anti-neoplastic prodrug
XX
PS Reference Example 14; Page 131; 182pp; English.
XX
CC A two-component system for anti-tumour therapy comprises a targeting
CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC prodrug. The system is based on antibody directed enzyme prodrug therapy
CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC (pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC be used. The present sequence encodes the native prepro-HCPB (see
CC AAW06172) contained in vector pIC11698.
XX
SQ Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;
XX
Alignment Scores:
Pred. No.: 3.75e-60 Length: 1263
Score: 637.50 Matches: 130
Percent Similarity: 54.85% Conservative: 68
Best Local Similarity: 36.01% Mismatches: 120
Query Match: 35.46% Indels: 43
DB: 17 Gaps: 6
US-09-980-881A-3 (1-338) x AAT42494 (1-1263)
QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
DB 58 TTTGAAGCGGAGAGGTGTTCCGTGTTAAGTTGAAGATGAAATACATTAACATAATC 117
QY 21 GlnAsnLeuThrThrThrThrGluLeuValLeuTrpGlnProValThrAlaAspLeu 40
DB 118 CGCGAGTTGGCCAGCAGCCAGATTGACTTCTTGGGAAGCCAGATTCTGTACACAAATC 177
QY 41 ValIysIysGlnValHisPhePheValAsnAlaSerAspValAspAsnValIysAla 60
DB 178 AAACCTCACAGTACAGTTGACTTCCGTTTAAAGCAGAGATATCTGCTGGGGAAT 237
QY 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 80
DB 238 GTTCTAAAGCAGATGAGTACATACAGTACTGATAGACCTGAGAAATGTGGTG 297
QY 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
DB 298 GAGGCTCAGTTGATACCGGGTT-----CGTCAACAGGACACAGTTATGAGAAG 348
QY 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
DB 349 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAGTCCGCACTGAGAAATCCAGCC 408
QY 121 MetLeuThrIleHisIleGlySerSerPheGluIysTyrProLeuTyrValLeuIys 140
DB 409 CTCATCTCTCGCAGTGTATCGGAACCAATTTGAGGGGCGCGCTATTATTCCTCTGGAAG 468
QY 141 ValSerGlyIysGluGlnThrAlaIysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
DB 469 GTT---GGCAAGCTGGACAAAATAAGCTGCCATTTTCATGACTGTGTGTTTCCATGCC 525
QY 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
DB 526 AGAGAGTGGATTTCTCTGCAATTCGCCAGTGGTTTGTAAAGAGAGGCTGTTCGTACCTAT 585
QY 173 ----- 173
DB 586 GGACGTGAGATCCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCTGCTGCTGTG 645
QY 174 -----GlyHis-----AsnArgMetTTPArgIysAsnArg 183
DB 646 CTCATATTATGATGGCTACATCTACACTGGACCAAGAGCCGATTTTGGAGAAAAGACTCGC 705
QY 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
DB 706 TCACCCATACTGGATCTAGCTGATTTGGCAGACAGACCCCAACAGAAATTTT---GATGCT 762
QY 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
DB 763 GGTGTGTGTAATTTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGACCTGCC 822
QY 224 ProGluSerGluProGluValIysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
DB 823 GCAGAGTCTGAAAGGAGACCAAGGCCCTGGCTGATTTTCATCGCAACAACTCTCTTCC 882
QY 244 IleValAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
DB 883 ATCAAGCATATCTGCAATCTCAATCTGATCTCCCAATGATGATCTACCTTACTCATAT 942
QY 264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
DB 943 GCTTACAAACTCGGTGAGAACATGCTGAGTTGAATGCCCTGCTAAAGCTACTGTGAAA 1002
QY 284 AlaIleGluIysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
DB 1003 GAACCTT---GCCTCACTGCAGCGCAAGTACACATATGCGCCGGAGCTACAACAATC 1059
QY 304 TyrLeuAlaProGlyGlyIysAspTrpIleTyrAspLeuGlyIleTyrSerPhe 323
DB 1060 TATCCTGTGCTGGGGGCTCTGACGACTGGGCTTATGACCAGGAATCAGATATTCCTTC 1119
QY 324 Thr 324
DB 1120 ACC 1122
RESULT 11
AAV41795
ID AAV41795 standard; DNA; 1263 BP.
XX
AC AAV41795;
XX
XX 20-NOV-1998 (first entry)
XX Human pancreatic carboxypeptidase nucleotide sequence.
DE
XX ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
XX prodrug therapy.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 1..1248
XX /tag= a
XX /product= "Pancreatic carboxypeptidase B"
XX /note= "No start codon given"
XX
XX WO9835988-A1.
XX
XX 20-AUG-1998.
XX
```



```
PF 10-FEB-1998; 98WO-GB00415.
XX
XX 29-OCT-1997; 97GB-0022727.
PR 14-FEB-1997; 97GB-0003104.
PR 18-OCT-1997; 97GB-0022003.
XX
XX (ZENE ) ZENEPA (ZENE ) ZENECA LTD.
XX
XX Edge MD;
XX
XX WPI; 1998-467168/40.
DR P-PSDS; AAW74476.
XX
XX New modified pro-domain of carboxy-peptidase B - enhances expression
PT of co-expressed proteins for production of recombinant
PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
XX enzyme prodrg therapy
XX
XX Example 1; Page 52-53; 83pp; English.
XX
XX The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC modified pro-domain of CPB on a separate gene to enhance recombinant
CC expression. This process can be used to produce recombinant CPB in
CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
CC insulin production and protein sequencing, while its fusions with
CC antibodies are useful in antibody-directed enzyme prodrg therapy. The
CC modified pro-domain provide increased yields of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.
XX
XX Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;
SQ

Alignment Scores:
Pred. NO.: 3.75e-60 Length: 1263
Score: 637.50 Matches: 130
Percent Similarity: 54.85% Conservative: 68
Best Local Similarity: 36.01% Mismatches: 120
Query Match: 35.46% Indels: 43
DB: 19 Gaps: 6

US-09-980-881A-3 (1-338) x AAV41795 (1-1263)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 58 TTTCAGCGGAGAGGTGTCCTCGTTACGTTGAAGTGAATCAATCAATCAATCAATC 117
Qy 21 GlnAsnLeuThrThrThrTyrGluLeValLeuTrpGlnProValThrAlaAspLeu 40
Db 118 CGCGAGTTGGCCAGCAGCCAGATTGACTTCTGGAAGCCAGATTCTGTGCACACAATC 177
Qy 41 VallyslyLysGlnValHisPhePheValAsnHisSerValAspValAspValLeuAla 60
Db 178 AAACCTCACAGTACAGTTGACTTCCTGTTAAACGAGAAATGATCTGCTGTGGAGAAT 237
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 80
Db 238 GTTCTTAAGCAGATGATCAATACAGTACTGATAGCAACCTGAGAAATGCTGTG 297
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 298 GAGGCTCAGTTTGATAGCGGGTT-----CGTGCAACAGGACACAGTTATGAGAAG 348
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 349 TACAACAAGTGGGAAACGATAGAGCTTGGACTCAACAGTGCCTGAGAAATCCAGCC 408
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 409 CTATCTCTCGCTGTTATCGGAACCCACATTTGAGGAGCGCGTATTATCTCTCTGAAG 468
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpLeuAspCysGlyIleHisAla 160
Db 469 GTT---GGCAAGCTGGCAAAATAAGCTGCCATTTTCATGGACTGTGGTTTCCATGCC 525
```

```
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
Db 526 AGAGAGTGGATTCTCTGCAATCTGCCAGTGGTTTGTAAAGAGAGGCTGTTGCTACCTAT 585
Qy 173 ----- 173
Db 586 GGACGTGAGATCCCAAGTGACAGAGCTTCTCGCAAGTTAGACTTTTATGTCTCTGCTGTG 645
Qy 174 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 183
Db 646 CTCAATATTGATGGCTACATCTACCTGGACCAAGCGGATTTTGGAGAAAGACTCGC 705
Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 706 TCACCCACTACTGATCTAGCTGATTTGGACAGAGCCCAACAGAAATTTT---GATGCT 762
Qy 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 763 GGTGGTGGTAATTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 822
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIleAsnGln 243
Db 823 GCAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGATTTCAATCGCAACAAACTCTCTCC 882
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 883 ATCAGGCATATCTGACATCCACTGTACTCCCAATGATGATCTACCTTACTCATAT 942
Qy 264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 943 GCTTACAAACTCGTGTGAGAAACAATGCTGAGTTCAATGCCCTGGCTAAAGCTACTGTGAAA 1002
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1003 GAACTT---GCCTCACTGCAGCCAGCAACAGTACACATATGCCCCGGGAGGTACACAATC 1059
Qy 304 TyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1060 TATCTGCTGCTGGGGCTCTGACACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1119
Qy 324 Thr 324
Db 1120 ACC 1122
RESULT 12
AAT42506
ID AAT42506 standard; DNA; 1284 BP.
XX
XX AAT42506;
XX
XX 17-FEB-1997 (first entry)
DE ProHCPB gene with PelB leader sequence.
XX
XX ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
XX mustard-ribonucleotide; antibody directed enzyme prodrg therapy;
XX anti-neoplastic; prodrg; reverse polarity; ion pair interaction;
XX reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
XX Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH CDS 1..1275 /*tag= a
FT FT 1..66 /*tag= b
FT sig_peptide /note= "PelB leader sequence"
FT sig_peptide 67..351 /*tag= c
FT FT /note= "pro sequence"
FT mat_peptide 352..1272
```


Db 1174 CCTTATGCATTGCTTTTCGAACTACGTGCACACTGGATATTTTGGATTTTACTCCGAG 1233
Qy 324 ----ThrSerAsnProProValGluLysLeu 332
Db 1234 ATGCTCATCAACCCACCCTGTACAGAACTA 1264

Search completed: January 1, 2004, 20:09:23
Job time : 391.048 secs

~~This Page Blank (suspect)~~

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:51:12 ; Search time 155.24 Seconds
(without alignments)
961.011 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FSGQVLAALPRTSRQVQL.....IKYFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09980881/runat_31122003_135917_26894/app.query.fasta_1.1628
-DB=Issued_Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 @CGN 1.115 @runat_31122003_135917_26894 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DSEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1749.5	97.3	1625	4	US-09-813-133A-1
2	1732	96.3	1272	2	US-08-869-057-1
3	1732	96.3	1749	1	US-07-649-591B-2
4	1732	96.3	1749	1	US-08-277-540-2
5	1732	96.3	1749	1	US-08-430-787A-2
6	646.5	36.0	1215	1	US-08-696-139-1
7	637.5	35.5	1263	2	US-08-860-882A-56
8	637.5	35.5	1263	4	US-09-011-769A-38
9	637.5	35.5	1284	2	US-08-860-882A-71
10	637.5	35.5	1284	4	US-09-011-769A-55
11	619.5	34.5	1311	4	US-09-675-305-9
12	619.5	34.5	2154	3	US-09-171-945-124

13	610	33.9	927	2	US-08-782-760-5	Sequence 5, Appli
14	610	33.9	927	5	PCT-US96-00995-5	Sequence 5, Appli
15	582	32.4	921	1	US-08-696-139-3	Sequence 3, Appli
16	571.5	31.8	2128	4	US-09-675-305-13	Sequence 13, Appli
17	567	31.5	999	2	US-08-860-882A-67	Sequence 67, Appli
18	567	31.5	999	4	US-09-011-769A-50	Sequence 50, Appli
19	567	31.5	1053	2	US-08-860-882A-64	Sequence 64, Appli
20	567	31.5	1053	4	US-09-011-769A-46	Sequence 46, Appli
21	560	31.1	1053	4	US-09-463-451-27	Sequence 27, Appli
22	560	31.1	1053	4	US-09-463-451-28	Sequence 28, Appli
23	560	31.1	1059	2	US-08-860-882A-74	Sequence 74, Appli
24	560	31.1	1059	4	US-09-011-769A-59	Sequence 59, Appli
25	559	31.1	1059	2	US-08-860-882A-77	Sequence 77, Appli
26	559	31.1	1059	4	US-09-011-769A-63	Sequence 63, Appli
27	554	30.8	1200	4	US-09-710-099-7	Sequence 7, Appli
28	549	30.5	1870	3	US-09-171-945-112	Sequence 112, App
29	534.5	29.7	1050	4	US-09-675-305-11	Sequence 11, Appli
30	527.5	29.3	1311	4	US-09-710-099-5	Sequence 5, Appli
31	516.5	28.7	1251	3	US-08-640-906-3	Sequence 3, Appli
32	516.5	28.7	1251	4	US-09-395-936-3	Sequence 3, Appli
33	491.5	27.3	1257	3	US-08-640-906-1	Sequence 1, Appli
34	491.5	27.3	1257	4	US-09-395-936-1	Sequence 1, Appli
35	467.5	26.0	945	4	US-09-710-099-3	Sequence 3, Appli
36	467.5	26.0	945	4	US-09-710-099-11	Sequence 11, Appli
37	465.5	25.9	2247	4	US-09-710-099-15	Sequence 15, Appli
38	441	24.5	1056	4	US-09-710-099-1	Sequence 1, Appli
39	441	24.5	1056	4	US-09-710-099-9	Sequence 9, Appli
40	365	20.3	55827	4	US-09-813-133A-3	Sequence 3, Appli
41	288.5	16.0	741	4	US-09-675-305-5	Sequence 5, Appli
42	210.5	11.7	591	3	US-09-331-709-2	Sequence 2, Appli
43	198.5	11.0	629	4	US-09-280-116-228	Sequence 228, App
44	169.5	9.4	494	4	US-09-669-751-63	Sequence 63, Appli
45	169.5	9.4	515	3	US-08-998-416-125	Sequence 125, App

ALIGNMENTS

RESULT 1
US-09-813-133A-1
; Sequence 1, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-1

Alignment Scores:
Pred. No.: 8,02e-210 Length: 1625
Score: 1749.50 Matches: 336
Percent Similarity: 94.38% Conservative: 0
Best Local Similarity: 94.38% Mismatches: 2
Query Match: 97.30% Indels: 18
DB: 4 Gaps: 1

US-09-980-881A-3 (1-338) x US-09-813-133A-1 (1-1625)

OY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
DB 83 TTTCAGATGGCCAGTTCTAGCTGCTCTCTCTAGAACCTCTAGGCAAGTTCAGTCTTA 142
OY 21 GlnAsnLeuThrThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuIle 40
|||||

Db 143 CAGAATCTTACTACACATATGAGATTGTTCTCTGCGAGCCGGTAACAGCTGACCTTATT 202
Qy 41 VallyyslyysGlnValHisPheValAsnAlaSerAspValAspValValLeuValLeu 60
Db 203 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTCATGTGCAATGTGAAGCC 262
Qy 61 HisLeuAsnValSerGlylleProCysSerValLeuLeuAlaAspValGluAspLeulle 80
Db 263 CATTTAAATGTGAGCGGAATCCATGCAGTGTCTGTCGACATGTGGAAGATCTTATT 322
Qy 81 GlnGlnGlnSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 323 CAACAGCAGATTTCCACAGCAGCAGTCCAGCCCGAGCCCTCCGATCGTACTATGAACAG 382
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 383 TATCACTCACTAAATGAATCTTCTTGTGATGAATTTATATCTGAGAGGCATCTCGAT 442
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 443 ATGCTTACAAATCCACATGATGATCTCTCATTTGAGNAGTACCACCTCATGTATTAAAG 502
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlylleHisAla 160
Db 503 GTTTCGAAAAAGAACAGCAGCAAAATGCCATATGGATTGACTGTGGAATCCATGCC 562
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
Db 563 AGAAGATGGATCTCTCTGCTTTCTGCTGTGTGTTTCATAGGCCCATATCGAATGTGAGA 622
Qy 181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
Db 623 AAGAACCGTCTTCTTATGGAACATCATGTGATCGGAACACCTGGAATAGAACCTTT 682
Qy 201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 220
Db 683 GCITCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGTCTCGGAACCTACTGT 742
Qy 221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
Db 743 GGAATTTATCTGATGATCAGAACCAAGAGTGAAGCAGTGGCTAGTTTCTTGAGAAGAAAT 802
Qy 241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
Db 803 ATCAACAGATTAAGCATATCATCATGATCATCTATCTCCAGCATATAGTTTCCA 862
Qy 261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
Db 863 TATTCCTATACAGAACTAAAGCAAGCAACCATGAGGAACCTGTCTCTAGTAGCCAGTGAA 922
Qy 281 AlavalArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
Db 923 GCAGTTCGTGTATTGAGAAAAATTAGTAAAAATACCAAGGTATACACATGGCCATGGCTCA 982
Qy 301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlylleLys 320
Db 983 GAAACCTTATACCTAGCTCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCATCAA 1042
Qy 321 TyrSerPhe----- 323
Db 1043 TATTCGTTTACAATTGAATTCGAGATACGGGCACATACGGATTCTTGTCTGCCGAGCGT 1102
Qy 324 -ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1103 TACATCAAAACCCCTCTAGAGAGCTTTTGGCGCTGTCTCTAAAA 1148

RESULT 2

US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko

; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA

; ZIP: 94804-0099

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/869,057

; FILING DATE: 03-JUN-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Washtien, Wendy L

; REGISTRATION NUMBER: 36,301

; REFERENCE/DOCKET NUMBER: 51509AUSM1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 510-262-5411

; TELEFAX: 510-262-7095

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1272 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; PUBLICATION INFORMATION:

; AUTHORS: Eaton, Dan L.

; AUTHORS: Malloy, Beth E.

; AUTHORS: Tsai, Siao P

; AUTHORS: Henzel, William

; AUTHORS: Drayna, Dennis

; TITLE: Isolation, Molecular Cloning, and Partial

; TITLE: Characterization of a No. 5985562el Carboxypeptidase B

; JOURNAL: J. Biol. Chem.

; VOLUME: 266

; ISSUE: 32

; PAGES: 21833-21838

; DATE: No. 5985562 15-1991

; US-08-869-057-1

Alignment Scores:

Pred. No.: 8,45e-208 Length: 1272
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-3 (1-338) x US-08-869-057-1 (1-1272)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 67 TTTCAAGTGGCCAAAGTTCTAGCTGCTCTCTCTAGAACCTCTAGGCAAGTTCAGTTCTA 126
Qy 21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeulle 40
Db 127 CAGAATCTTACTACAAACATATGAGATTGTTCTCTGCGAGCCGGTAACAGCTGACCTTATT 186
Qy 41 VallyyslyysGlnValHisPheValAsnAlaSerAspValAspValValLeuAla 60
Db 187 GTGAAGAAAAACAAGTCCATTTTGTAAATGCATCTCATGTGCAATGTGAAGCC 246
Qy 61 HisLeuAsnValSerGlylleProCysSerValLeuLeuAlaAspValGluAspLeulle 80

Db 247 CATTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGCAGACGCTGGAAGATCTTAAT 306
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 307 CAACACAGATTTTCCAACGACAGATGACAGCCCGAGCCTCCGCACTGCTACTATGACAG 366
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 367 TATCACTCACTAATGAATCTATTCTTGATAGAATTTATAACTGAGAGCATCTGTAT 426
Qy 121 MetLeuThrIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 427 ATGCTTACAAAATCCACATGATCTCTCATTTGAGAAGTACCCTACTATGTTTAAAG 486
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 487 GTTTCGGAAGAACAAACAGCAAAATGCCATATGGATTGACGTGGAATCCATGCC 546
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
Db 547 AGAAGATGGAATCTCTCTGCTTCTGCTGTGTGTTCTATAGGCCATATAACTCAATCTAT 606
Qy 175 ----- 175
Db 607 GGGATAATAGGCAATATACCAATCTCTGAGGCTTGTGATTTCTATGTTATGCCGTG 666
Qy 176 -----AsnArgMetTrpArgLysAsnArg 183
Db 667 GTTAATGTGGAGCGTTATGACTACTCATGCAAAAGCAATCGAATGTGGAGAAAGACCGT 726
Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 727 TCTTTCTATGCGAACAATCATTCATCGGAACAGACTGAATAGGAACCTTGTCTCCAA 786
Qy 204 HisTrpCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 787 CACTGTGTGAGAAGGTGATCCAGTCTCTCATGTCTGGAACCTACTGTGGACTTTAT 846
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIleAsnGln 243
Db 847 CCTGAGTCAGAACCAAGTGAAGGCGAGTGGCTAGTTCTTGAGAAGAAATATCAACAG 906
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 907 ATTAAGCATATACATGATGATTCATATCTCCAGCATATAGTGTTCATATTCCTAT 966
Qy 264 ThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 967 ACAGGAATGAAGCAAGCAACCAATGAGCACTGTCTAGTACCCAGTGAAGCGTTCGT 1026
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1027 GCTATTGAGAAACTAGTAAATAATACAGGTATACACATGGCCATGGCTCAGAAACCTTA 1086
Qy 304 TyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1087 TACCTAGCTCCTGGAGGTGGGAGCATGTGATCTATGATTTGGGCATCAATATTCGTT 1146
Qy 324 -----ThrSerAs 326
Db 1147 ACAATTGAATCTCGATACGGGCACATACGGATTTCTTGCTCGGAGCGTTACATCAA 1206
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1207 CCCACCTGTAGAGAAGCTTTGCGCGTGTCTCTAATA 1243

RESULT 3

US-07-649-591B-2

; Sequence 2, Application US/07649591B

; Patent No. 5206161

; GENERAL INFORMATION:

; APPLICANT: Dennis Drayna and Daniel Eaton

; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/649,591B
FILING DATE: 19910201
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Haak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/266-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: hybridization probe
LOCATION: 133 to 178
IDENTIFICATION METHOD:
OTHER INFORMATION:
FEATURE:
NAME/KEY: potential clip site
LOCATION: 380 to 382
IDENTIFICATION METHOD:
OTHER INFORMATION:
FEATURE:
NAME/KEY: signal sequence
LOCATION: 41 to 106
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-07-649-591B-2

Alignment Scores:
Pred. No.: 1,44e-207 Length: 1749
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 1 Gaps: 2

US-09-980-881A-3 (1-338) x US-07-649-591B-2 (1-1749)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 107 TTTCAGATGGCCAGATTCCTAGCTCTCTTAGAACCTCTAGGCAAGTTCAAGTTCTA 166
Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuIle 40
Db 167 CAGATCTTACTACACATATGAGATTGTTCTCGCAGCCGCTAACAGCTGACCTTAAT 226
Qy 41 ValLysLysGlnValHisPhePheValAsnAlaSerAspValAspValLysAla 60
Db 227 GTGAAGAAAAACCAAGTCCATTTTTTTGTAATGTCATCTGATGTCGACCAATGTGAAGCC 286

QY 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLeu 80
DB 287 CATTTAAATGTGAGCGAATTCATGAGTGTCTGTCGACAGCGTGGAAGATCTATT 346
QY 81 GlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
DB 347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCGAGCCCTCCGCATCGTACTATGAACAG 406
QY 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
DB 407 TATCACTCACTAAATGAATCTATTCTTGGATAGAAATTTATATACTGAGAGSCATCTGAT 466
QY 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
DB 467 ATGCTTACAAAATCCACATGATCTCTATTGAGAGTACCACCTCTAATGTTTAAAG 526
QY 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
DB 527 GTTCTTGAAAGAACAAACAGCCAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586
QY 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
DB 587 AGAAGATGGATCTCTCTGCTTTCTGCTGTGTTTCATAGGCCATATAACTCAATTCTAT 646
QY 175 ----- 175
DB 647 GGGATAATAGGCAATATACCAATCTCTGAGCGTTGTGGAATTTCTATGTTATGCCGGTG 706
QY 176 -----AsnArgMetTrpArgLysAsnArg 183
DB 707 GTTAAATGTGACGGTTATGACTACTCATGCAAAAAGAAATCGAATGTGGAGAAAGACCGT 766
QY 184 SerPheTyrAlaSerHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
DB 767 TCTTTCTATCGAACATCATTCATCGAACAGACCTGAATAGGAACCTTTGCTTCCAA 826
QY 204 HisTrpCysGluGluGlyAlaSerSerCysSerGluThrTyrCysGlyLeuTyr 223
DB 827 CATGGTGTGAGGAGTGATCCAGTTCTCTCATGCTCGGAAACCTACTGTGGACTTAT 886
QY 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
DB 887 CCTGAGTCAGAACCAAGTGAAGGAGTGGCTAGTTCTTGAGAGAAATATCAACCAG 946
QY 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
DB 947 ATTAAAGCATATCATCATGATGCTATTCATCTCCAGCATATAGTGTTCATATTCCTAT 1006
QY 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
DB 1007 ACACGAAGTAAAGCAAGACCAATGAGAACTGTCTAGTACCCAGTGGAAGCTTCGT 1066
QY 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
DB 1067 GCTATTGAGAAACTAGTAAATATACCAAGTATACATGCGCATGCGCTCAGAAACCTTA 1126
QY 304 TyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
DB 1127 TACCTAGCTCCTGAGGTGGGACGATTTGATCTATGATTGGGCATCAATATTCGTT 1186
QY 324 -----ThrSerAs 326
DB 1187 ACAATTGAATTCGAGATACGGGCACATACGGAATTTCTGTCGGGAGCGTTATACATAA 1246
QY 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
DB 1247 CCCACCTGTAGAGAGCTTTGCGCGTGTCTCTAANA 1283

RESULT 4

US-08-277-540-2

; Sequence 2, Application US/08277540

; Patent No. 5474901

; GENERAL INFORMATION:

; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1CID1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-277-540-2

Alignment Scores:
Pred. No.: 1,44e-207 Length: 1749
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-3 (1-338) x US-08-277-540-2 (1-1749)

QY 1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
DB 107 TTTTCAGATGGCCCAAGTCTTAGCTGCTCTTCTTAGAACCTCTAGGCAAGTTCAAGTCTTA 166
QY 21 GlnAsnLeuThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuLeu 40
DB 167 CAGAACTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATT 226
QY 41 ValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
DB 227 GTCAAGAAAAACAGTCCCATTTTGTAAATGCATCTGATGTCGACATGTGAAAGCC 286
QY 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLeu 80
DB 287 CATTTAAATGTGAGCGGAATTCATGAGTGTCTTCTGGCAGCGGTGAAGATCTTATT 346
QY 81 GlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
DB 347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCGAGCCCTCCGCATCGTACTATGAACAG 406

Qy 101 TyrHisSerLeuAenGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 407 TATCACTACATAAATGAATCTATCTTGGATAGATTTTATACTGAGAGCATCTCGAT 466
Qy 121 MetLeuThrLysIleHisIleGlySerPheGluLysTyrProLeuTyrValLeuLys 140
Db 467 ATGCTTACAAAATCCATGATGATCTCTATTGAGAGTACCCACTTAATGTTTAAAG 526
Qy 141 ValSerGlyLysGluInThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 527 GTTCTCGAAAAGAACAAACAGCCAAAATATGCCATATGGATTGACTGTGAATCCATGCC 586
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
Db 587 AGAAGATGGATCTCTCTGCTTCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTAT 646
Qy 175 ----- 175
Db 647 GGGATAATAGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706
Qy 176 -----AsnArgMetTrpArgLysAsnArg 183
Db 707 GTTAATGTGGACGGTTATGACTACTCATCGAAAAGAAATCGAATGTGAGAAACCGT 766
Qy 184 SerPheTyrAlaAenAsnHisCysIleGlyThrAspLeuAenArgAsnPheAlaSerLys 203
Db 767 TCTTTCTATGCGAACATCATTTGATCGAACACAGACCTGAATAGGAACCTTGTCTTCAAA 826
Qy 204 HistPrcGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 827 CACTGGTGTGAGGAAGTGATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTTAT 886
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 887 CCTGAGTCAGAACAGAGTGAAGCAGTGGCTAGTTCTTGAGAAAGAAATCAACACAG 946
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 947 ATTAAGCATACATCAGCATGCTATCTATCCAGCATATAGTGTTCATATTCCTAT 1006
Qy 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 1007 ACAGGAAGTAAAGAACAGACCATGAGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1067 GCTATTGAGAAACTAGTAAATATACAGGTATACATATGGCCATGGCTCAGAAACCTTA 1126
Qy 304 TyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1127 TACCTAGCTCTGAGGCTGGGACGATTTGGATCTATGATTTGGGCATCAAAATATTCGTT 1186
Qy 324 -----ThrSerAs 326
Db 1187 ACAATTGAATCTCGAGATACGGGCACATACCGATTTCTGTCGGGAGCGCTTATCAAA 1246
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1247 CCCACCTGTAGAGAGCTTTTGGCGCTGTCTTAAAA 1283

RESULT 5

US-08-430-787A-2
; Sequence 2, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco

STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-430-787A-2

Alignment Scores:
Pred. No.: 1,44e-207 Length: 1749
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservatives: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 1 Gaps: 2

US-09-980-881A-3 (1-338) x US-08-430-787A-2 (1-1749)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 107 TTTTCAGAGTGGCCAAAGTTCTAGCTGCTTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166
Qy 21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
Db 167 CAGAACTTTACTCAACATATGAGATTGTCTCTGCGAGCGCGGTAAACAGCTTATTT 225
Qy 41 ValLysLysLysGlnValHisPheValAsnAlaSerAspValAspValLysAla 60
Db 227 GTGAAGAAAAACAAGTCCATTTTTTTGTAATATGCATCTCATGTGCAATATGTAAGGCC 286
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
Db 287 CATTTAAATGTGAGCGGAATTCATGCAGTGTCTTGTGCGACAGCTGGAGATCTTATTT 346
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 347 CAACAGCAGATTTTCAACAGCACAGTCCAGCCCGAGCCCTCCGATCTGCTATGAACAG 406
Qy 101 TyrHisSerLeuAenGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 407 TATCACTACATAAATGAATCTATCTTGGATAGATTTTATACTGAGAGCATCTCGAT 466

Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 467 ATGCTTACAAAATCCACATGGATCTCTCAATTCAGAGATACCACTCTATGTTTAAAG 526
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrrPleAspCysGlyIleHisAla 160
Db 527 GTTCTGGAAGAAGAACAAACAGCCAAATAATGCCATATGGATTGACTGTGGAATCCATGCC 586
Qy 161 ArgGluTrrPleSerProAlaPheCysLeuTrrPheIleGlyHis 175
Db 587 AGAATGGATCTCTCTGCTTTCTCTGTTGTTTCATAGGCCATATAAATCAATTCAT 646
Qy 175 ----- 175
Db 647 GGGATATAGGCCATATACCAATCTCTGAGGCTTGTGGATTCTATGTTATGCCGGTG 706
Qy 176 -----AsnArgMetTrrPArgLysAsnArg 183
Db 707 GTTAATGTGACGGTTATGACTACTCATCGAAGAAAGAAATCGAATGTGGAGAAACCGT 766
Qy 184 SerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 767 TCTTTCTATCGCAACATCATGTGATCGAACACCTGAATAGGAATTTGCTTCCAA 826
Qy 204 HisTrrPLeuGluGlyAlaSerSerCysSerGluThrTrrCysGlyLeuTyr 223
Db 827 CACTGGTGTGAGGAGGTGATCCAGTTCCTCATGCTCGGAACCTACTGTGGACTTAT 886
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 887 CCTGAGTCAGAACAGAGAGTGAAGGAGTGGCTAGTTCTTGAAGAAATATCAACAG 946
Qy 244 IleLysAlaTrrLleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 947 ATTAAGCATACATCAGCATGCTATCTACTCCAGCATATAGTGTTCATATTCCTAT 1006
Qy 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
Db 1007 ACAGAGATTAAGCAAGACCATGAGAACTGCTCTTAGTAGCCAGTGAAGCAATTCGT 1066
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTrrThrHisGlyHisGlySerGluThrLeu 303
Db 1067 GCTATTGAGAAACTAGTAAAAATACCAGGTATACATGCGCATGCGCTCAGAAACCTTA 1126
Qy 304 TyrLeuAlaProGlyGlyLysAspTrrPleTrrAspLeuGlyIleLysTyrSerPhe 323
Db 1127 TACCTAGCTCTGAGGTGGGACGATTTGGATCTATGATTTGGGCATCAATATTCGTTT 1186
Qy 324 -----ThrSerAs 326
Db 1187 ACAATGACTTCAGATACGGGCACATACGAGTCTTGTCGCGGAGCGTTATCATCAA 1246
Qy 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db 1247 CCCACCTGTAGAGAAGCTTTTGCGCGTCTCTTAA 1283

RESULT 6

US-08-696-139-1

; Sequence 1, Application US/08696139

; Patent No. 5672496

; GENERAL INFORMATION:

; APPLICANT: Fayerman, Jeffrey T.

; APPLICANT: Greenen, David P.

; APPLICANT: Herhsberger, Charles L.

; APPLICANT: Larson, Jeffrey L.

; APPLICANT: Sterner, Jane L.

; APPLICANT: Zhang, Haichao

; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Eli Lilly and Company

; STREET: Lilly Corporate Center

; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/696,139
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1215 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1215
; US-08-696-139-1

Alignment Scores:

Pred. No.: 2,96e-71 Length: 1215
Score: 646.50 Matches: 137
Percent Similarity: 54.57% Conservative: 60
Best Local Similarity: 37.95% Mismatches: 121
Query Match: 35.96% Indels: 43
DB: 1 Gaps: 6

US-09-980-881A-3 (1-338) x US-08-696-139-1 (1-1215)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 28 TTCGAAGGGGAGAGGTTTCCGTCATGTTTGAAGATGAAATGACATCAGCTTACT 87
Qy 21 GlnAsnLeuThrThrTrrTyrGluIleValLeuTrrPLeuProValThrAlaAspLeuIle 40
Db 88 CATGAGTTGGCCAGCACCGCAGATTGACTTCTGGAACCCAGATCTGTGCACAAATC 147
Qy 41 VallysLysLysGlnValHisPheValAsnAlaSerAspValAsnValLysAla 60
Db 148 AAACCTCACAGTACAGTTGACTTCCGTTGGAAGCAGACATATTTGGCTGTGGAAGAC 207
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
Db 208 TTTCTGGAGCAGAAATGAATACAAATATGAGGTACTATAAACCACTGAGATCTGTGCTC 267
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTrrGluGln 100
Db 268 GAGCTCAGTTTGCAGCAGATC-----CGTCAACTGGACACAGTTATGAGAAG 318
Qy 101 TyrHisSerLeuAsnGluIleTrrSerTrrPLeuPheIleThrGluArgHisProAsp 120
Db 319 TACAACAACCTGGGAACGATCGAGGCTTGGACTAAGCAAGTACCAGTGAATAATCCAGAC 378
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 379 CTCATCTCTCGCACAGCCATCGGAACACTACATATTTTAGGAACAATATATATACCTCTCAAG 438

Db 763 GGTGGTGTGAATTTGAGCGCTCTCGAAACCCCTGTGTGATGAAACTTACTGTGGACCTGCC 822

Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnGln 243
|||||
Db 823 GCAGAGTCTGAAAGGAAACCAAGCGCTGTGATTTTCATCCGCAACAAACTCTCTCC 882
|||||
Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
|||||
Db 883 ATCAGGCATATCTGCAATCCACTCGTACTCCCAATGATGATCTACCCCTTACTCATAT 942
|||||
Qy 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaLaserGluAlaValArg 283
|||||
Db 943 GCTTACAACTCGGTGAGAACAACTGCTGAGTTGAATCCCTCGGTAAAGTACTGTGAAA 1002
|||||
Qy 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
|||||
Db 1003 GAACCTT---GCCTCCTGCTGCGGACCAAGTACATATGCGCGGAGCTACAACTC 1059
|||||
Qy 304 TyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
|||||
Db 1060 TATCCTGCTGCTGGGGCTCTGACGACTGGCTTATGACCAAGGAATCAGATATTCCTTC 1119
|||||

Qy 324 Thr 324
|||

Db 1120 ACC 1122

RESULT 9

US-08-860-882A-71

; Sequence 71, Application US/08860882A

; Patent No. 5985281

; GENERAL INFORMATION:

; APPLICANT: TAYLORSON, CHRISTOPHER JOHN

; APPLICANT: EGGELTE, HENDRIKUS JOHANNES

; APPLICANT: TARRAGONA-PIOL, ANTONIO

; APPLICANT: RABIN, BRIAN ROBERT

; APPLICANT: BOYLE, FRANCIS THOMAS

; APPLICANT: HENNAM, JOHN FREDERICK

; APPLICANT: BLAKELY, DAVID CHARLES

; APPLICANT: MARSHAM, PETER ROBERT

; APPLICANT: HEATON, DAVID WILLIAM

; APPLICANT: DAVIES, DAVID HUW

; TITLE OF INVENTION: CHEMICAL COMPOUNDS

; NUMBER OF SEQUENCES: 77

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PILLSBURY, MADISON & SUTRO

; STREET: 1100 NEW YORK AVENUE, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy Disk

; COMPUTER: IBM compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/860,882A

; FILING DATE: JUNE 23, 1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: DONALD J. BIRD

; REGISTRATION NUMBER: 25,323

; REFERENCE/DOCKET NUMBER: 9901/238653

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3027

; TELEFAX: (202) 822-0944

; TELEX: 6174627 CUSH

; INFORMATION FOR SEQ ID NO: 71:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1284 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-860-882A-71

Alignment Scores: 4,41e-70 Length: 1284

Pred. No.: 637,50 Matches: 130

Percent Similarity: 54,85% Conservative: 68

Best Local Similarity: 36,01% Mismatches: 120

Query Match: 35,46% Indels: 43

DB: 2 Gaps: 6

US-09-980-881A-3 (1-338) x US-08-860-882A-71 (1-1284)

Qy 1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
|||||

Db 85 TTTGAAGCGGAGAGGTGTTCCGTGTTAACGTTGAAGATGAAATCAATTAACATAATC 144
|||||

Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuIle 40
|||||

Db 145 CGCGAGTTGGCCAGCAGCAGCCAGATTGACTTCTGGNAGCCAGATTCTGTGCACAAATC 204
|||||

Qy 41 ValLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60
|||||

Db 205 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAGACTGTCTGCTGGAGAAAT 264
|||||

Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
|||||

Db 265 GTTCTAAAGCAGAAATGAACATAACAGTACTGATAAGCAACCTGAGAAATGTGGTG 324
|||||

Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
|||||

Db 325 GAGGCTCAGTTGATAGCCGGTT-----CGTGCAACAGGACACAGTTATGAGAA 375
|||||

Qy 101 TyrHisSerLeuAsnGluIleTyrSerTyrPheIleGluPheIleThrGluArgHisProAsp 120
|||||

Db 376 TACAACAAGTGGGAAACGATAGAGCTTGGACTCAACAGCTGCCACTGAGAAATCCAGCC 435
|||||

Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
|||||

Db 436 CTCATCTCTCGCAGTGTATTCGGAACCAACATTTGAGGGACGCGCTATTTACTCTCTGAAG 495
|||||

Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrPheIleAspCysGlyIleHisAla 160
|||||

Db 496 GTT---GGCAAGCTGGACAAAATAAGCTGCCATTTTCATGAGACTGTGTGTTCCATGCC 552
|||||

Qy 161 ArgGluTyrPheSerProAlaPheCysLeuTyrPheIle----- 173
|||||

Db 553 AGAGAGTGGATTTCTCTCGATTTCTGCCACTGTGTTGTAAGAGAGGCTGTTCGTACCTAT 612
|||||

Qy 173 ----- 173

Db 613 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCTCTGCTGTG 672
|||||

Qy 174 -----GlyHis-----AsnArgMetTyrArgLysAsnArg 183
|||||

Db 673 CTCATATTGATGGCTACATCTACACTGGACCAAGAGCGGATTTTGGAGAAAGACTCGC 732
|||||

Qy 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
|||||

Db 733 TCCACCCATACTGGATCTAGCTGATGGACAGACCCCAACAGAAATTTT---GATGCT 789
|||||

Qy 204 HisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
|||||

Db 790 GGTGGTGTGAAATTTGGAGCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 849
|||||

Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
|||||

Db 850 GCAGAGTCTGAAAGGAGACCAAGGCCCTGGCTGATTTTCATCCGCAACAACTCTCTTCC 909
|||||

Qy 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
|||||

Db 910 ATCAGGCATATCTGCAATCCACTCGTACTCCCAATGATGATGTACTACCTTACTCATAT 969
|||||

Qy 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAlaValArg 283
|||||

Db 1030 GAACCTT---GCCTCACTGCACGGCACCAGTACACATATATGGCCGGGAGCTACAAATC 1086
Qy 304 TyLeuAlaProGlyGlyAspAspTrpIleTyAspLeuGlyIleTySerPhe 323
Db 1087 TATCCTGCTGCTGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATCTCTC 1146
Qy 324 Thr 324
Db 1147 ACC 1149
RESULT 11
US-09-675-305-9
; Sequence 9, Application US/09675305
; Patent No. 644153
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 644153el Human Carboxypeptidases and
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/09/675,305
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: US 60/156,685
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-675-305-9
Alignment Scores:
Pred. No.: 8.4e-68 Length: 1311
Score: 619.50 Matches: 136
Percent Similarity: 51.15% Conservative: 64
Best Local Similarity: 34.78% Mismatches: 127
Query Match: 34.45% Indels: 64
DB: 4 Gaps: 8
US-09-980-881A-3 (1-338) x US-09-675-305-9 (1-1311)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 109 TATGCTGGTGATAAAGTAAAGTAAAGTAAATTTATTTCCCAAAACAGAGGAGGATATGCACTG 168
Qy 21 GlnAsnLeuThrThrThrTyGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
Db 169 AAGAAATATCCTATCACTCACTTAAAGTGGAGCTGTGGCGCCAGCAGTATCTCTATGTA 228
Qy 41 ValLys-----LysLysGlnValHisPhePheValAsnAlaSerAspValAspVal 58
Db 229 TCAGAGGGAACAGTACTGATGTCATATATCCCAAAATGTTCCCGAGCC-----CTG 282
Qy 59 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 78
Db 283 TTAGCCCTTCTTACAGGAAGCAACATCCAGTACAAAGGTCTCTATAGAAAGATCTTCAGAA 342
Qy 79 LeuIleGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSer----- 96
Db 343 AACTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
Qy 97 Ty-TyrGluGlnTyHisSerLeuAsnGluIleTySerTrpIleGluPheIleThrGlu 116
Db 403 AATTATCAAGTTTATCACTCTCTAGAGAAATTCAAATTTGGATGTCATCTGAATAAA 462
Qy 117 ArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyProLeu 136
Db 463 ACTCACTCAGGCTCTATCACTGTTCTCTATTGGAGATCATATGAGGAGGAGGAGGAGGAGGAG 522

Qy 137 TyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCys 156
Db 523 TTTATTTTAAAGCTG---GGCAGACGATCAGCACTCAAAAGAGCTGTTGGATAGACTGT 579
Qy 157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
Db 580 GGTATTTCATGCAAGAGATGGATTGGCTCTGCTCTTTTGTGAGTGGTTTGTAAAGAGACT 639
Qy 173 ----- 173
Db 640 CTTCTAATATATAAGAGTACCAGCCATGATGAGAAATTTGTAATCATCTATATTTCTAT 699
Qy 174 -----GlyHisAsnArgMetTrp 179
Db 700 ATCATGCTGTGTTTAAACGTTCGATGATACCATTTTGTAGTGGACCAATGATCGATTGG 759
Qy 180 ArgLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 199
Db 760 AGAAAAACAAGGTCAAGGAACCAAGTTTCGCTGCGGTGGAGTGGATGCCAATAGAAAC 819
Qy 200 PheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTy 219
Db 820 TCGAAGTGAAG---TGGTGTGATGAGGAGCTTCTATGCACCTTGTGATGACACATAC 876
Qy 220 CysGlyLeuTyProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
Db 877 TGTGGCCCTTTTCCAGAACTGAGCGGAAGTGAAGCTGTAGCTTAACCTTCTTCGAAA 936
Qy 240 AsnIleAsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisIleValPhe 259
Db 937 CACAGAAAGCACATAGGGCTTATCTCTCTTTCATGCATATGCTCAGATGTTACTGTAT 996
Qy 260 ProTySerTyThrArgSerLysAspHisGluGluLeuSerLeuValAlaSer 279
Db 997 CCTATCTTACAAATATGCAACATTTCCCAATTTAGATGTGTGAATCTGCAGCTTAT 1056
Qy 280 GluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisGlyHisGly 299
Db 1057 AAAGCTGTGAATGCACTTTCAGTCAGTA---TACGGGTACGATACAGATATGACCAAGCC 1113
Qy 300 SerGluThrLeuTyLeuAlaProGlyGlyAspTrpIleTyAspLeuGlyIle 319
Db 1114 TCCCAACGTTGTATGTAGCTCTGCTAGCTCAATGATGGGCTTACAAAAATGGAATA 1173
Qy 320 LysTySerPhe----- 323
Db 1174 CCTATGCAATTTGCTTCCAACTACGTGACACTGGATATTTGGATTTTACTCCAGAG 1233
Qy 324 ---ThrSerAsnProValGluLysLeu 332
Db 1234 ATGCTCATCAAAACCCACCTGTACAGAACTA 1264
RESULT 12
US-09-171-945-124
; Sequence 124, Application US/09171945
; Patent No. 6277599
; GENERAL INFORMATION:
; APPLICANT: Emery, Stephen
; APPLICANT: Copley, Clive Graham
; APPLICANT: Edge, Michael Derek
; TITLE OF INVENTION: Monoclonal Antibody to CEAs
; TITLE OF INVENTION: Antibody, and Their Therapeutic Use in an Adept System
; FILE REFERENCE: Monoclonal Antibody to CEAs
; CURRENT APPLICATION NUMBER: US/09/171,945
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: GB9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; NUMBER OF SEQ ID NOS: 131

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
; LENGTH: 2154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: humanized
; US-09-171-945-124

Alignment Scores:
Pred. No.: 1 94e-67 Length: 2154
Score: 619.50 Matches: 127
Percent Similarity: 54.29% Conservative: 69
Best Local Similarity: 35.18% Mismatches: 122
Query Match: 34.45% Indels: 43
DB: 3 Gaps: 6

US-09-980-881A-3 (1-338) x US-09-171-945-124 (1-2154)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 64 TTTGAAGGCGGAGAGGTGTTCCGTGTTAAAGTGAAGATGAAATACACATTAAACAATC 123
Qy 21 GlnAsnLeuThrThrThrTyrGluLeuValLeuThrGlnProValThrAlaAspLeuLe 40
Db 124 CGCGAGTTGCCGACGACGACCCAGATTGCTTCTCGAAGCCAGATTCTGTCCACACAATC 183
Qy 41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspValLysAla 60
Db 184 AAACCTCACAGTACAGTGTGACTTCGCTGTTAAAGCAGAGATGCTGCTCACTGTGGAGAT 243
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLe 80
Db 244 GTTCTAAAGCAGAGTAACACTACATACAGGTACTGATGAAGCAACCTGAGAAATGTGGTG 303
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTy-TyrGluGln 100
Db 304 GAGGCTCAGTTTGATAGCGGGTT-----CGTGAACACAGGACACAGATTATGAGAAG 354
Qy 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 355 TACAACAAGTGGGAACGATAGAGGCTTGATCTCAACAGTGCCTCCACCTGAGAAATCCAGCC 414
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db 415 CTCACTCTCGCAGTGTATCGGACCACTTTGAGGACGCGCTATTACTCTCTGAAG 474
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 475 GTT--GGCAAAAGCTGGACAAATAAGCTGCCATTTTCATGGACTGTGGTTTCCATGCC 531
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
Db 532 AGAGAGTGGATTCTCTCGCATTTCTGCCAGTGGTTTGTAAAGAGAGGCTGTTTCGTACCTAT 591
Qy 173 ----- 173
Db 592 GGACGTGAGATCCAAAGTACAGAGCTTCTCGAAGATTAGACTTTTATGTCCTCGCTGTG 651
Qy 174 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 183
Db 652 CTCAATATTGATGCTACATCTACACCTGGACCAAGAGCGGATTTTGGAGAAAGACTGCG 711
Qy 184 SerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 712 TCCACCCATAGTGGATCTAGCTGTCATTGGCAGACAGACCCCAACAGAAATTTT---GATGCT 768
Qy 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 769 GGTGGTGGTGAATTTGAGGCTCTCGAAACCCCTGTGATGAACATTTACTTGGAGCTGCC 828
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIleAsnGln 243
Db ----- 243

```

Alignment Scores:

Pred. No.: 7,35e-67 Length: 927
Score: 610.00 Matches: 126
Percent Similarity: 60.50% Conservative: 44
Best Local Similarity: 44.84% Mismatches: 71
Query Match: 33.93% Indels: 41
DB: 2 Gaps: 5

US-09-980-881A-3 (1-338) x US-08-782-760-5 (1-927)

QY 93 AlaserAlaserTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 112
DB 1 GCAAGTGGACACAGCTACACCAAGTACCAAGTGGGAACGATTGAGCGGTGGATTCAA 60
QY 113 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 132
DB 61 CAAGTTGCCACTGATAATCCAGACCTTGTCCTCAGAGCGTCAATGGAACACCATTTGAA 120
QY 133 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 152
DB 121 GGAGCTAAACATGATGTCTCAAGATT--GGTAAACTAGACCGCAATAGCGCTGCCATC 177
QY 153 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 172
DB 178 TTCATCGATTGGTTTCCATGCAAGAGAGTGATTCTCTCGATCTCTCAGTGGTTT 237
QY 173 Ile----- 173
DB 238 GTGAGAGAGCGTCTCGTACCTATATCAAGAGATCCACATGAACAGCTTCTAGATGAA 297
QY 174 -----GlyHis----- 175
DB 298 CTGGATTTCTATGTTCTGCTGTGCTGCAACATTCATGCTATGCTACACCTGGCAATAG 357
QY 176 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 195
DB 358 GACAGATGTGGAGAAACCCGCTCTACTGCTGGAAAGTTCCTGCTTGGGTGTAGAC 417
QY 196 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 215
DB 418 CCCAACAGGAATTT---AATGCTGGCTGTGTGAAGTGGAGCTTCTCGAGTCCCTGC 474
QY 216 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 235
DB 475 TCTGAACCTTACTGTGACCCAGAGCTCTGAAAGAGAGACAAAGGCGCTGGCAGAT 534
QY 236 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 255
DB 535 TTCATCGCAACACCTCTCCACATCAAGGCTTACTGACCATCCATCACTACTACAG 594
QY 256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 275
DB 595 ATGATGCTCTACCTTACTCTATGACTACAACTGCCTGAGAACTATGAGGAATGAAT 654
QY 276 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 295
DB 655 GCCCTGTGAAAGTGGCGGCAAGGAGCTT---GCCACTCTGATGGCACCAGATGACACA 711
QY 296 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyr 315
DB 712 TATGGCCCGAGGAGCTACAAATCTATCTCTGCTGGGGGATCTGACGAGCTGGTCTAT 771
QY 316 AspLeuGlyIleLysTyrSerPheThrSerAsnProValGluLysLeuLeuProLeu 335
DB 772 GATCAGGAATCAATATTCCTTACCTTTGAACT--CCGGGATACAGGCTTCTTGGCTT 830
QY 336 Ser 336
DB 831 TCT 833

RESULT 14

PCT-US96-00995-5

; Sequence 5, Application PC/TUS9600995
; GENERAL INFORMATION:
; APPLICANT: Bio-Technology General Corp.
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00995
; FILING DATE: 25-JAN-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..927
; PCT-US96-00995-5

Alignment Scores:

Pred. No.: 7,35e-67 Length: 927
Score: 610.00 Matches: 126
Percent Similarity: 60.50% Conservative: 44
Best Local Similarity: 44.84% Mismatches: 71
Query Match: 33.93% Indels: 41
DB: 5 Gaps: 5

US-09-980-881A-3 (1-338) x PCT-US96-00995-5 (1-927)

QY 93 AlaserAlaserTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 112
DB 1 GCAAGTGGACACAGCTACACCAAGTACCAAGTGGGAACGATTGAGCGGTGGATTCAA 60
QY 113 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 132
DB 61 CAAGTTGCCACTGATAATCCAGACCTTGTCCTCAGAGCGTCAATGGAACACCATTTGAA 120
QY 133 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 152
DB 121 GGAGCTAAACATGATGTCTCAAGATT--GGTAAACTAGACCGCAATAGCGCTGCCATC 177
QY 153 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 172
DB 178 TTCATCGATTGGTTTCCATGCAAGAGAGTGATTCTCTCGATCTCTCAGTGGTTT 237
QY 173 Ile----- 173
DB 238 GTGAGAGAGCGTCTCGTACCTATATCAAGAGATCCACATGAACAGCTTCTAGATGAA 297

174 -----GlyHis----- 175
298 CTGGAATTTCTATGTTCTGCTGTGGTCAACATGATGCTATGCTACACCTGGACTAG 357
176 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 195
358 GACAGATGTGGAGAAACCCGCTCTACTATGCTGGAAAGTTCCTGCTGGGTGTAGAC 417
196 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGluAlaSerSerSerCys 215
418 CCCAAGAGGAATTT---AATGCTGGCTGTGGTGAAGTGGAGCTTCTCGAGTCCCTGC 474
216 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 235
475 TCTGAATCTTACTGTGGACAGCCCGAGAGTCTGAAAGAGACAAAGGCCCTGGAGAT 534
236 PheLeuArgArgAsnIleAlaGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 255
535 TTCAATCGCAACCACTCTCCACCATCAAGGCTACCTGACCATCCACTCATACTCACAG 594
256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluLeuSer 275
595 ATGATGCTCTACCTTACTCTCTACTACTACAACTGCTGAGAACTATGAGGAATTGAAT 654
276 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 295
655 GCCCTGTGAAGTGGCGCAAGGAGCTT---GCCACTGTGATGGCACAAGTACACA 711
296 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 315
712 TATGCGCCAGGAGCTACCAACAATCTATCTGCTGCTGGGGATCTGACGAGTGTCTTAT 771
316 AspLeuGlyIleLysTyrSerPheThrSerAsnProValGluLysLeuLeuProLeu 335
772 GATCAGGAATCAATATCTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830
336 Ser 336
831 TCT 833

RESULT 15
US-08-696-139-3
Sequence 3, Application US/08696139
Patent No. 5672496
GENERAL INFORMATION:
APPLICANT: Fayerman, Jeffrey T.
APPLICANT: Greenen, David P.
APPLICANT: Hersberger, Charles L.
APPLICANT: Larson, Jeffrey L.
APPLICANT: Sterner, Jane L.
APPLICANT: Zhang, Haichao
TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
OPERATING SYSTEM: PC-DOS/MS-DOS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,139
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,258
FILING DATE: 16-NOV-1993

ATTORNEY/AGENT INFORMATION:
NAME: Gavlo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..921
US-08-696-139-3

Alignment Scores:
Pred. No.: 2,42e-63 Length: 921
Score: 582.00 Matches: 117
Percent Similarity: 59.47% Conservative: 40
Best Local Similarity: 44.32% Mismatches: 67
Query Match: 32.37% Indels: 40
DB: 5 Gaps: 5

US-09-980-881A-3 (1-338) x US-08-696-139-3 (1-921)

QY 98 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 117
DB 16 TATGAGAAGTACAACTCGGGAACGATCGAGGCTTGGACTAAGCAAGTCACCAAGTAA 75
QY 118 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 137
DB 76 AATCCAGAGCTCATCTCTGCGACAGCCATCGGAACATACATTTTATAGGAACAATATATAC 135
QY 138 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 157
DB 136 CTCCTCAAGGTT---GGCAACCTGGACCAATAAGCTGCCATTTTCATGGAGCTGTGT 192
QY 158 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
DB 193 TTCCATGCCAGAGATGGATTTCCCATGCATTTTGGCAGTGGTGTGTGAGAGAGGCTGTT 252
QY 173 ----- 173
DB 253 CTCACCTATGGATATGAGAGTCACATGACAGAAATCTCAACAAGCTAGACTTTTATGTC 312
QY 174 -----GlyHis-----AsnArgMetTrpArg 180
DB 313 TTGCTGTGCTCAATATGATGGCTATACCTGACCTGGACCAAGAACCGAATGTGGAGA 372
QY 181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
DB 373 AAGACCGCTCTACCAATGCTGGAACCTACTGATTTGGCAGACACCCACAGAAATTT 432
QY 201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 220
DB 433 ---GATGCTGGGTGGTGCACACTGGAGGCTCTACAGACCCCTGCGACTTACTGT 489
QY 221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 240
DB 490 GGATCTGCTGCAGAGTCTGAAAGAGACCAAGGCCCTGGCTGATTTTATACCAACAAC 549
QY 241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
DB 550 CTCTCTCCATCAAGCATACCTGACGATCCACTCATCTACAGATGATACTTACCCT 609
QY 261 TyrSerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaSerGlu 280
DB 610 TATTCTCTATGATTACAAATCCCCGAGAACAAATGCTGAGTTGAATAAATCCTGGCTAAGCT 669

```
QY 281 AlavAlaAlaAlaGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
Db 670 GCCGTGAAGACTT---GCTACTGTATGGCACCAGTACACATACGGCCCGAGGCT 726
QY 301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
Db 727 ACAACAATCTATCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAA 786
QY 321 TyrSerPheThr 324
Db 787 TATTCCTTCACC 798
```

Search completed: January 2, 2004, 09:25:54
Job time : 177.24 secs

OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-980-107-2396

Alignment Scores:

Pred. No.: 5.72e-211 Length: 1728
Score: 1732.00 Matches: 338
Percent Similarity: 86.01% Conservative: 0
Best Local Similarity: 86.01% Mismatches: 0
Query Match: 96.33% Indels: 55
DB: 10 Gaps: 2

US-09-980-881A-3 (1-338) x US-09-880-107-2396 (1-1728)

```
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 86 TTTCCAGAGTGGCCAAAGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTCTA 145
Qy 21 GlnAsnLeuThrThrThrTyrrGluileValLeuTrpGlnProValThrAlaAspLeuile 40
Db 146 CAGAATCTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATT 205
Qy 41 VallyslylsGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
Db 206 GTGAAGAAAAACAAGTCCATTTTTTTGTAATGCATCTGATGTCGACAATGTGAAGCC 265
Qy 61 HistLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuile 80
Db 266 CATTTAAATGTAGCGGAATTCATGACGATGCTTCTGCGCAGCGGTAAACAGCTGACCTTATT 325
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrrTyrrGluGln 100
Db 326 CAACAGCAGATTTCCACAGCAGACAGTACGCCCCGAGCCTCCGATCGTACTATGAACAG 385
Qy 101 TyrHisSerLeuAsnGluileTyrrSerTrpIleGluPheleThrGluAArgHisProAsp 120
Db 386 TATCACTCACTAAATGAATCTATTCTTGGATGAATTTATATACTGAGAGCGATCTCGAT 445
Qy 121 MetLeuThrIleHisIleGlySerSerPheGlyIleTyrrProLeuTyrrValLeuLys 140
Db 446 ATGCTTACAAAATCCCAATGGATCTCTCATTTGAGAAGTACCCACCTCTATGTTTTAAAG 505
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 506 GTTCTCGAAAGAAACAAACAGCCAAAATGCCATATGGATTGACTGTGGAAATCCATGCC 565
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis----- 175
Db 566 AGAATGGATCTCTCTGCTTCTGCTGTGGTTTCATAGGCCATATAAATCAATTCAT 625
Qy 175 ----- 175
Db 626 GGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 685
Qy 176 -----AsnArgMetTrpArgLysAsnArg 183
Db 686 GTTAATGTGACGGTTATGACTACTATCGTGAAGAAAGATCGATGTGGAAGAACCGT 745
Qy 184 SerPheTyrrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db 746 TCTTTCTATGCGAACATCATTTGCATCGGAACAGACCTGGAATAGAACTTTGCTTCCAAA 805
Qy 204 HistTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrrCysGlyLeuTyrr 223
Db 806 CACTGGTGTGAGGAAGTGATCCAGTTCTCATGTCGGAACCTACTGTGGACTTTAT 865
Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db 866 CTTGAGTCAGAACAGAGTGAAGCGAGTGGCTAGTTCTTGAAGAAATATCAACAG 925
Qy 244 IleLysAlaTyrrIleSerMetHisSerTyrrSerGlnHisIleValPheProTyrrSerTyrr 263
Db 926 ATTAAGCATACATCAGCATGCTTACTATCCAGCATATAGTGTGTTTCCATATTCCTAT 985
```

RESULT 2

US-10-379-836-1

; Sequence 1, Application US/10379836

; Publication No. US20030215850A1

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON

; FILE REFERENCE: D0214NP

; CURRENT APPLICATION NUMBER: US/10/379, 836

; CURRENT FILING DATE: 2003-03-04

; PRIOR APPLICATION NUMBER: U.S. 60/361,523

; PRIOR FILING DATE: 2002-03-04

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent in version 3.2

; SEQ ID NO 1

; LENGTH: 1272

; TYPE: DNA

; ORGANISM: Papio hamadryas

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1269)

US-10-379-836-1

Alignment Scores:

Pred. No.: 7.23e-198 Length: 1272
Score: 1628.00 Matches: 318
Percent Similarity: 83.21% Conservative: 9
Best Local Similarity: 80.92% Mismatches: 11
Query Match: 90.55% Indels: 55
DB: 13 Gaps: 2

US-09-980-881A-3 (1-338) x US-10-379-836-1 (1-1272)

```
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 67 TTTCCAGAGTGGCCAGGTTCTAGCTGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTCTA 126
Qy 21 GlnAsnLeuThrThrThrTyrrGluileValLeuTrpGlnProValThrAlaAspLeuile 40
Db 127 CAGAATCTTACTACACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCGGACCTTATT 186
Qy 41 VallyslylsGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
Db 187 GAGAAGAAAAACAAGTCCATTTTTTTGTAATTCATCTGATGTCGACAATGTGAAGCC 246
Qy 61 HistLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuile 80
Db 247 CATTTAAATGTGAGCGAATTCATGACGATGCTCTGCGCAGATGGAAGATCTTATT 306
Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrrTyrrGluGln 100
Db 307 CAACAGCAGATTTCCAAACGACAGTACGAGCCCGAGCCTCCGATCGTACTACTGAACAG 366
```

QY 101 TyrHisSerLeuAnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
DB 367 TATCACTCACTAAATGAATCTATCTTGGATAGAACTTATACTGAGAAGTATCTCGAT 426
QY 121 MetLeuThrIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
DB 427 ATCTTACAAATCCACATGGATCTCTCTGATGAGAACCCCACTTATGTTTAAAG 486
QY 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
DB 487 GTTCTTGGAAAGAACAAACAGCCAAATGATGCGATGGATTGACGTGGAAATCCATGCC 546
QY 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
DB 547 AGAGAATGGATCTCCCTGCTTCTGCTGTGTGTTTCATAGGCCATATAACTGAATACTAC 606
QY 175 ----- 175
DB 607 GGGATAATAGGGGAATATACCAATCTTCTGAGGCATGTGGATTTCTATGTTATGCCAGTG 666
QY 176 -----AsnArgMetTrpArgLysAsnArg 183
DB 667 GTTAATGTGGATGTTTATGACTACTCATGGAAGAAAGATCGAATGTGGAGAAGACCGT 726
QY 184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
DB 727 TCTTTCTATGCGAACAATCTGTCATCGGAACAGACCTGAACAGGAACCTTTGCGTCCAA 786
QY 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
DB 787 CACTGTGTGAGGAAGGTGCATCCAGTTCTCATGCTCGGAACCTACTGTGGACTTTAT 846
QY 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
DB 847 CTTGAGTCAGAACAGAGGAGGCGGTGCTAAATTTCTTGAGAAGAAATATCAACCAC 906
QY 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
DB 907 ATTAAGCATACATCAGCATTCATCTCCAGCATATCGTTTCCATATTCCTAT 966
QY 264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
DB 967 ACTCGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1026
QY 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
DB 1027 GCTATTTCAGAAACCAAGTAAATATATCAGGTATACATGCGCGTGGCTCAGAAACCTTA 1086
QY 304 TyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
DB 1087 TACCTAGCTCCTGGAGTGGGAGCATTTGGATCTATGATTTGGGCATCAATATTCGTTT 1146
QY 324 -----ThrSerAs 326
DB 1147 ACAATTGAATCTCGAGATACGGCAATACGGATTTCTTGCTGCTGAGCGTTATACAAA 1206
QY 326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
DB 1207 CCACCTTGTAAGACCGCTTTTGGCGTGTCTCTAAAA 1243

RESULT 3

US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (25)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Alignment Scores:

Prod. No.: 8.81e-105 Length: 1400
Score: 902.00 Matches: 181
Percent Similarity: 76.05% Conservative: 0
Best Local Similarity: 76.05% Mismatches: 1
Query Match: 50.17% Indels: 56
DB: 9 Gaps: 2

US-09-980-881A-3 (1-338) x US-09-925-302-24 (1-1400)

QY 157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
DB 10 GGAATCCATGCCAGANAATGGATCTCTCTGCTTTCTGCTGTGTTCTATAGGCCATATA 69
QY 175 ----- 175
DB 70 ACTCAATTTCTATGGGTAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTAT 129
QY 176 -----AsnArgMetTr 179
DB 130 GTTATGCCGGTGGTTAATCTGGATGCTTATGAATCTCATGGAAGAAAGATCGAATGTG 189
QY 179 pArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAs 199
DB 190 GAGAAAGAACCGTTCTTTCTATCGGAACAATCATTCGATCGGAACAGACCTGAATAGGA 249
QY 199 nPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTy 219
DB 250 CTTTGTCTTCAACACTGTGTGTGAGGAAGGTGCATCCAGTTCTCTATGCTCGGAACCTTA 309
QY 219 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 239
DB 310 CTGTGCACTTTATCTGAGTCAGAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 369
QY 239 gAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 259
DB 370 AAATATCAACCAAGATTAAGCATACATCAGCATGCATTCATCTCCAGCATATAGTGTT 429
QY 259 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 279
DB 430 TCCATATTCCTATACAGGAAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 489
QY 279 rGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 299
DB 490 TGNAGCAGTTCGTGCTATTGAGANAACACTAGTAAANAATACAGGTATACATCGCCATGG 549
QY 299 ySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIle 319
DB 550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGAGCGATTGGATCTATGATTTGGGCAT 609
QY 319 elysTyrSerPhe----- 323
DB 610 CAAATATTCGTTTACAAATTGAATTCGAGATACGGGCACATACGGATTTCTGTGCCGA 669
QY 324 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLys 338
DB 670 GCGTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAAA 719

RESULT 4

```
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299. A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: ME101-1561RNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; PRIOR FILING DATE: 2002-08-28
; PRIOR FILING DATE: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-229-546-3
Alignment Scores:
Pred. No.: 5,48e-69 Length: 1254
Score: 622.00 Matches: 128
Percent Similarity: 55.12% Conservativeness: 71
Best Local Similarity: 35.46% Mismatches: 118
Query Match: 34.59% Indels: 44
DB: 7 Gaps: 7

US-09-980-881A-3 (1-338) x US-10-229-546-3 (1-1254)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 61 TTGACAGGAGAGGTGTTCCGGTGAAGCCAGGATGAAACAAAGCAGCATCAT 120
Qy 21 GlnAsnLeuThrThrThrThrGluLeuValLeuTrpGlnProValThrAlaAspLeu 40
Db 121 AAGGACTTGGCCAAACCAATAGCTTGTGATCCAGGTGTCAGGTCGCCACCACCGTA 180
Qy 41 VallyslylsGlnValHisPheValAsnAlaSerAspValAspAsnValysAla 60
Db 181 GCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAGGAATCCCAAGCCATCCAGTCT 240
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 80
Db 241 GCCTTGATCAATAAATGCAATGCAATCTTGATTCATGATCTACAGAGAGATT 300
Qy 81 GlnGlnGlnLeuSer---AsnAspThrValSerProArgAlaSerAlaSerTyThrGlu 99
Db 301 GAGAAACAGTTGATGTTAAAGAGATATCCAGGAGGAGGACAGC-----TAGCA 351
Qy 100 GlnTyHisSerLeuAsnGluIleTySerTrpIleGluPheIleThrGluArgHisPro 119
Db 352 AAATACATAATTTGGGAAAGATTGTGCTTGGACTGAAAGATGATGATGATATATCCT 411
Qy 120 AspMetLeuThrIleHisIleGlySerSerPheGlySerValProLeuTyThrValLeu 139
Db 412 GAATGCTCTCTGATTAATAATTTGATCTACTGTTGAAGATAATCCACTATATGTTCTG 471
Qy 140 LysValSerGlyIysGluInThrAlaIleAsnAlaIleTrpIleAspCysGlyIleHis 159
Db 472 AAGATT---GGGAAAAAGAAATGAAAGAAAGCTATTTTATGGATTGTCGATTCAC 528
Qy 160 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
Db 529 GCACGAGATGGGTCTCCCGCAGCATTTGCGAGGTGTTGTCTATCAGGCAACCAAACT 588
Qy 173 ----- 173

US-10-229-802-21
; Sequence 21, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF) -RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-21
Alignment Scores:
Pred. No.: 8,43e-69 Length: 1622
Score: 622.00 Matches: 128
Percent Similarity: 55.12% Conservativeness: 71
Best Local Similarity: 35.46% Mismatches: 118
Query Match: 34.59% Indels: 44
DB: 7 Gaps: 7

US-09-980-881A-3 (1-338) x US-10-429-802-21 (1-1622)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 61 TTGACAGGAGAGGTGTTCCGGTGAAGCCAGGATGAAACAAAGCAGCATCAT 120

589 TATGGGAGAAACAAAATATATGACCAAACTCTTGGACCGAATGAATTTTACATCTTCTCT 648
174 -----GlyHis-----AsnArgMetTTDAAGLYAsn 182
649 GTGTTCAATGTTGATGATATATTTGGTCATGGACAAAGACCGCATGTGGAGAAAAAT 708
183 ArgSerPheTyThrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
709 CGTTCAGAACCAAACTCCAATGCTGCGACTGACCTCAACAGGAATTTT---AAT 765
203 LysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyThrCysGlyLeu 222
766 GCTTCAGAACCTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGCTCT 825
223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
826 GCACAGAGTCCGAGAAAGAGCTGCTCACTAATTTCAATTAAGAAGCCACTGAAT 885
243 GlnIleLysAlaTyTrpIleSerMetHisSerTyThrGlnHisIleValPheProTyThr 262
886 GAAATCAAGTTTACATCACCTTCCATTCCTCTCCAGATGCTATGTTTCCCTATGGA 945
263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
946 TATACATCAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAAGATTGGCACT 1005
283 ArgAlaIleGluLysThrSerLysAsnThrArgTyThrHisGlyHisGlySerGluThr 302
1006 GATGTTCTTA---TCAACTCGATATGAACCCGCTACATCTATGCGCCCAATAGAAATCAACA 1062
303 LeuTyLeuAlaProGlyGlyGlyAspTrpIleTyAspLeuGlyIleLysTyThrSer 322
1063 ATTACCGATATCAGGTTCTTCTTAGACTGGGCTTATGACCTGGGCATCAAAACACACA 1122
323 Phe 323
1123 TTT 1125

RESULT 5
US-10-429-802-21
; Sequence 21, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF) -RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-21
Alignment Scores:
Pred. No.: 8,43e-69 Length: 1622
Score: 622.00 Matches: 128
Percent Similarity: 55.12% Conservativeness: 71
Best Local Similarity: 35.46% Mismatches: 118
Query Match: 34.59% Indels: 44
DB: 7 Gaps: 7

US-09-980-881A-3 (1-338) x US-10-429-802-21 (1-1622)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 61 TTGACAGGAGAGGTGTTCCGGTGAAGCCAGGATGAAACAAAGCAGCATCAT 120
```


589 TATGGGAGAAACAAATTTATGACCAAACTCTTGGACCGAATGAATTTTACATTCTTCCT 648
174 -----GlyHis-----AsnArgMetTyrArgLysAsn 182
649 GTGTTCAATGTGTGATGATATATTGTCATGGACAAAGAACCGCATGTGGAGAAAAAT 708
183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
709 CGTTCGAAGAACCAAACTCCAATGATGCTGGCACTGGCACTCAACAGGAATTTT---AAT 765
203 LysHisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeu 222
766 GCTTCATGGAATCCATTCCTAACCAACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825
223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIleAsn 242
826 GCACGAGTCCGAGAGAGACGAAAGCTGTCACTAATTTCAATAGAACCCACTGAAT 885
243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 262
886 GAATCAAGGTTTACATCACTCCATTCCTACTCCAGATGCTATTGTTTCCCTATGGA 945
263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
946 TATACATCAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTCGCAAGATTGGCACT 1005
283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
1006 GATGTTCTTA---TCAACTCGATATGAACCGCTACATCATCTATGGCCCAATAGAATCAACA 1062
303 LeuTyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSer 322
1063 ATTTACCGATATCAGGTTCTCTTTAGACTGGGCTTATGACCTGGGCATCAACACACA 1122
323 Phe 323
1123 TTT 1125

RESULT 7
US-10-341-434-187
; Sequence 187, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341.434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
; LENGTH: 1633
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)..(1262)
; OTHER INFORMATION:
US-10-341-434-187

Alignment Scores:
Pred. No.: 8.53e-69 Length: 1633
Score: 622.00 Matches: 128
Percent Similarity: 55.12% Conservative: 71
Best Local Similarity: 35.46% Mismatches: 118
Query Match: 34.59% Indels: 44
DB: 13 Gaps: 7

US-09-980-881A-3 (1-338) x US-10-341-434-187 (1-1633)
Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 72 TTTGACGAGGAGAAGGTGTTCCGCGTGAAGCCCGCAGGATGAACAAACAGCAGCATCAT 131
Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuThrGlnProValThrAlaAspLeuIle 40
Db 132 AAGGACTTGGCCAAACCAATGAGCTTGCTTGGTATCCAGGTGCCACCCACCGTA 191
Qy 41 ValLysLysGlnValHisPhePheValAsnAlaSerAspValAspValLysAla 60
Db 192 GCTGCTAATATGATGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCAGCCAGTCT 251
Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
Db 252 GCCTTGGATCAAAATAAATGCACTATGAATCTTGAATCATGATCTACAAGAAGAGATT 311
Qy 81 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 99
Db 312 GAGAACAGTTTGATGTTAAAGAGATATCCAGGCGAGGCACAGC-----TAGCCA 362
Qy 100 GlnTyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArgHisPro 119
Db 363 AAATCAATAAATTCGCAAAAGATTGTGGCTTGGACTGAAAGATGATGGAATAAGTATCCT 422
Qy 120 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 139
Db 423 GAAATGGTCTCGTATTAATAATTTGGATCTACTGTTGAAGATTAATCCACTATATGTTCTG 482
Qy 140 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrIleAspCysGlyIleHis 159
Db 483 AAGATT---GGGAAAAGATGAAGNAGNAGAGGCTATTTTATGATGTGGCATCAC 539
Qy 160 AlaArgGluTyrIleSerProAlaPheCysLeuTyrPheIle----- 173
Db 540 GCACGAGAATGGTCTCCCGCATTTCTGCAGTGTGTTGTCCTATCAGGCAACCAAACT 599
Qy 173 ----- 173
Db 600 TATGGGAGAAACAAATTTATGACCAAACTCTTGGACCGAATGAATTTTACATTCTTCCT 659
Qy 174 -----GlyHis-----AsnArgMetTyrArgLysAsn 182
Db 660 GTGTTCAATGTGTGATGATATATTGTCATGGACAAAGAACCGCATGTGGAGAAAAAT 719
Qy 183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
Db 720 CGTTCGAAGAACCAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 776
Qy 203 LysHisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeu 222
Db 777 GCTTCATGGAATCCCATTCCTAACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 836
Qy 223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
Db 837 GCACGAGTCCGAGAGAGAGAGAGAGAGCTGTCACTAATTTTATTAGAACCCACTGAAT 896
Qy 243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 262
Db 897 GAAATCAAGGTTTACATCACTTCCATTCCTACTCCAGATGCTATTGTTTCCCTATGGA 956
Qy 263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
Db 957 TATACATCAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTCGCAAGATTGGCACT 1016
Qy 283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
Db 1017 GATGTTCTA---TCAACTCGATATGAACCGCTACATCTATGGCCCAATAGAATCAACA 1073
Qy 303 LeuTyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSer 322
Db 1074 ATTTACCGGATATCAGGTTCTCTTTAGACTGGGCTTATGACCTGGGCATCAACACACA 1133

QY 323 Phe 323
 Db 1134 TTT 1136

RESULT 8
 US-10-200-344-9
 ; Sequence 9, Application US/10200344
 ; Publication No. US20020173641A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Nehls, Michael
 ; APPLICANT: Friedrich, Glenn
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.
 ; TITLE OF INVENTION: No. US20020173641A1 Human Carboxypeptidases and
 ; FILE OF INVENTION: Polynucleotides Encoding the Same
 ; FILE REFERENCE: LEX-0047-USA
 ; CURRENT APPLICATION NUMBER: US/10/200,344
 ; CURRENT FILING DATE: 2002-07-19
 ; PRIOR APPLICATION NUMBER: US/09/675,305
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 1311
 ; TYPE: DNA
 ; ORGANISM: homo sapiens
 US-10-200-344-9

Alignment Scores:
 Pred. No.: 1,23e-68 Length: 1311
 Score: 619.50 Matches: 136
 Percent Similarity: 51.15% Conservative: 64
 Best Local Similarity: 34.78% Mismatches: 127
 Query Match: 34.45% Indels: 64
 DB: 14 Gaps: 8

US-09-980-881A-3 (1-338) x US-10-200-344-9 (1-1311)

QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 Db 109 TATGCTGGTGAATAAGATGAAGATTATTTCCCAAAAGAGAGGAGCATATGCACTG 168
 QY 21 GlnAsnLeuThrThrThrTyGluLeValLeuTyrGlnProValThrAlaAspLeu 40
 Db 169 AAGAAATATCCTATCACTTAAGTGGACCTGTGGCAGCCGACGATCTCTATGTA 228
 QY 41 ValLys-----LysLysGlnValHisPheValAsnAlaSerAspValAspVal 58
 Db 229 TCAGAGGGAACAGTTACTGATGTCATATCCCCCAAAATGGTCCCGAGCC-----CTG 282
 QY 59 LysAlaHisLeuAsnValSerGlyLeProCysSerValLeuLeuAlaAspValGluAsp 78
 Db 283 TTAGCCTCTTTCAGGAGGAGCAACATCCAGTACAGGTCCTCATAGAAGATCTTCAGAA 342
 QY 79 LeuileGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSer----- 96
 Db 343 AACTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
 QY 97 TyrTyGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 116
 Db 403 AATTATGAAGTTTATCACTCTTAGAGAAATTCANAATGGATGCATCATCTGAATAA 462
 QY 117 ArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeu 136
 Db 463 ACTCACTCAGGCGCTCATTCATCTCTATTGGAAGATCATATGAGGAGGAGGAGGAGGAG 522
 QY 137 TyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrPheAspCys 156
 Db 523 TTTATTTTAAAGCTG---GGCAGACGATCACTCAAGCTCAAAAGAGCTGTTGGATAGACTGT 579

QY 157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
 Db 580 GGTATTTCATGCAAGAGAATGGATTGGTCTGCTCTTTTGTGAGTGTGTGAAAAAGAGCT 639
 QY 173 ----- 173
 Db 640 CTTCTAACATATAAGAGTGACCCAGCCATGAGAAATGTTGAATCATCTATATTTCTAT 699
 QY 174 -----GlyHisAsnArgMetTrp 179
 Db 700 ATCATGCTGTGTTTAACTGCGATGATACCAATTTTAGTTGGACCAATGATGATTTGG 759
 QY 180 ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 199
 Db 760 AGAAAAACAAGGTCAGGAACCTCAAGGTTTCGCTGCCGTGAGTGCATGCAATAGAAC 819
 QY 200 PheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyr 219
 Db 820 TGGAAAGTGAAG---TGGTGTGATGAAGAGGCTTCTATGACCCCTGTGTGATGACATAC 876
 QY 220 CysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 239
 Db 877 TGTGGCCCTTTTCCAGAACTGAGCCGGAAGTGAAGCTGTAGCTAATCTCTTCGAAAA 936
 QY 240 AsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhe 259
 Db 937 CACAGAAAGCACATTAGGGCTTATCTCTCTCTTCATGATGCTCAGATGTTACTGTAT 996
 QY 260 ProTyrSerTyrThrArgSerLysSerLysAspHisGluLysLeuValAlaSer 279
 Db 997 CCTATTCTTACAAATATGCAAAATCCCAATTTTAGATGTGTGGAATCTGCAGCTTAT 1056
 QY 280 GluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGly 299
 Db 1057 AAAGCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGACCCAGCC 1113
 QY 300 SerGluThrLeuTyrLeuAlaProGlyGlyAspTrpIleTyrAspLeuGlyIle 319
 Db 1114 TCACAACGTTGTATGTAGCTCTGCTGATGCTCAATGATTTGGGCTTACAAAAATGGAATA 1173
 QY 320 LysTyrSerPhe----- 323
 Db 1174 CTTATGCAATTTGCTTTCGAACCTACGTCAGCTGGATATTTTGGATTTTACTCCAGAG 1233
 QY 324 ---ThrSerAsnProValGluLysLeu 332
 Db 1234 ATGCTCATCAACCCACCTGTACAGAACTA 1264

RESULT 9
 US-10-274-639-33
 ; Sequence 33, Application US/10274639
 ; Publication No. US20030232349A1
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE GENOMICS, INC.
 ; APPLICANT: DESLEGEAN, Angelo M.; GANDHI, Ameena R.
 ; APPLICANT: HAFALIA, April J.A.; LU, Dyung Aina M.
 ; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
 ; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
 ; APPLICANT: NGUYEN, Daniel B.; LEE, Ernestine A.
 ; APPLICANT: KHAN, Farrah A.; YOS, Henry
 ; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
 ; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
 ; APPLICANT: YANG, Junming; THANGAVELU, Kavitha
 ; APPLICANT: DING, Li; KEARNEY, Liam
 ; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
 ; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
 ; APPLICANT: BURFORD, Neil; WALIA, Narinder K.
 ; APPLICANT: LAL, Preeti G.; LEE, Sally
 ; APPLICANT: TODD, Stephen; LO, Terence P.
 ; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
 ; APPLICANT: AZIMZAI, Yalda; LU, Yan
 ; TITLE OF INVENTION: PROTEASES
 ; FILE REFERENCE: PI-0167 USA

207 AAACCTCACAGTACAGTTGACTTCCGGTAAAGCAGAGATACCTGCTGCTGGAGAT 266
 61 HisLeuValSerGlyLeuProCysSerValLeuLeuAlaSerValGluAlaSerLeu 80
 267 GTTCTAAGCAGAGTAACATAACAGTACTGATAGCAACCTGAGAAATGGTGG 326
 81 GluGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 327 GAGCTCAGTTGATAGCCGGTT-----CGTCAACAGCAGACAGTTATGAGAG 377
 101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
 378 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCCCACTGAGATCCAGCC 437
 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
 438 CTCATCTCTCGAGTGTATTCGAACACACATTTGAGGAGCGCGCTATTATACCTCTGGAAG 497
 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
 498 GTT---GGCAAGCTGGACAAATAAGCTGCCATTTCATGAGCTGTGGTTTCCATGCC 554
 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 173
 555 AGAGAGTGGATTCTCTCGATTCTGCGAGTGTGTGTAAGAGAGGCTGTCTGCTACCTAT 614
 173 ----- 173
 615 GGAGCTGAGATCCAAAGTGCAGAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCTGTG 674
 174 -----GlyHis-----AsnArgMetTrpArgLysAsnArg 183
 675 CTCATATTGATGCTACATCTACCTCGACCAAGAGCCGATTTTGGAGAAAGACTCGC 734
 184 SerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
 735 TCCACCATACTGATCT---AGCATGGCAGACAGCCCAACAGAAATTT---GATGCT 788
 204 HisTrpCysGluGluAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
 789 GGTGGTGTGAAATGGAGGCTCTCGAAACCCCTGTGATGAAACTTACTGTGACCTGCC 848
 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsnIleAsnGln 243
 849 GCAGAGTCTGAAAGGAGACCAAGGCTGTGCTGATTTCACTCCGCAACAACTCTCTTC 908
 244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
 909 ATCAAGGCATATCTGACAACTCCATCTCTCCAAATGATGATCTACCTTACTCATAT 968
 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAlaValArg 283
 969 GCTTACAACTCGGTGAGAACAAATGCTGAGTTGAATGCCCTGCTAAAGCTACTGTGAA 1028
 284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
 1029 GAACCT---GCCATCTGCGCAGCAGCAAGTACATATGGCCGGGAGCTACAACTC 1085
 304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
 1086 TATCTCTGCTGGGGCTCTGAGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1145
 324 Thr 324
 1146 ACC 1148

RESULT 12

US-09-960-352-14595
 ; Sequence 14595, Application US/09960352
 ; Patent No. US20020137139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing
 ; APPLICANT: Byatt, John C.
 ; APPLICANT: Mathialagan, Nagappan
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 ; FILE REFERENCE: 16511.006/37-21 (10298)C
 ; CURRENT APPLICATION NUMBER: US/09/960,352
 ; CURRENT FILING DATE: 2001-09-24
 ; NUMBER OF SEQ ID NOS: 15112
 ; SEQ ID NO 14595
 ; LENGTH: 416
 ; TYPE: DNA
 ; ORGANISM: Bos taurus
 ; OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10
 US-09-960-352-14595

Alignment Scores: 8,12e-66 Length: 416
 Pred. No.: 591.00 Matches: 111
 Score: 89.13% Conservative: 12
 Percent Similarity: 80.43% Mismatches: 15
 Best Local Similarity: 32.87% Indels: 0
 Query Match: 10 Gaps: 0
 DB:

US-09-980-881A-3 (1-338) x US-09-960-352-14595 (1-416)

QY 29 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysGlnValHisPhe 48
 DB 2 ATTGTCTCTGCGAGCCAGCTAGCAGCTGAATATATTGTGAAGGATACCAAGTCCATTTT 61
 QY 49 PheValAsnAlaSerAspValAsnValLysAlaHisLeuAsnValSerGlyIlePro 68
 DB 62 TTTGTGAATGTCATCTGATGTAGCAATGTGAAGCCCATTTAATCGGAGCAGAAATCCA 121
 QY 69 CysSerValLeuLeuAlaSerValGluAspLeuIleGlnGlnIleSerAsnAspThr 88
 DB 122 TTCAGGGTCTCTGGTGAATAATGTGAAGATCTTATCCGGCAGCAGACTTCCAATGACACC 181
 QY 89 ValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyr 108
 DB 182 ATCAGCCCCCGGCGATCTCTCTCTACTATGACAGTATCACTCACTAAATGAGATCTAT 241
 QY 109 SerTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 128
 DB 242 TCTTGATAGAGTTATGACTGACGCGTATCTGATATGTTGAAAAAATCCACATTTGA 301
 QY 129 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAla 148
 DB 302 TCCTCATACAGAGTACCACCTTTATGTTTAAAGGTTTCTAAAAAGAACCAAGGGCC 361
 QY 149 LysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 166
 DB 362 AAAAATGCCATGTGGATTGACTGTGGAATCCAGCCAGAGTGGATCTCTCTCT 415

RESULT 13

US-10-200-344-13
 ; Sequence 13, Application US/10200344
 ; Publication No. US20020173641A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Nehls, Michael
 ; APPLICANT: Friedrich, Glenn
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.
 ; TITLE OF INVENTION: Polynucleotides Encoding the Same
 ; FILE REFERENCE: LEX-0047-USA
 ; CURRENT APPLICATION NUMBER: US/10/200,344
 ; CURRENT FILING DATE: 2002-07-19
 ; PRIOR APPLICATION NUMBER: US/09/675,305
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 13


```
Db      835 GTTCTTAACATAGATGTTATATCTACACTTGGACAACATGATCGTCTTTGGAGGAATCC 894
Qy      183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
Db      895 CGTTACCCCATATAATGGCACATGTTTGGACGGATCTCAATCGAAATTTC--AAT 951
Qy      203 LysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeu 222
Db      952 GCATCTTGGTGTAGTATGTTGGCTCTAGAACTGCCAAGATCAACAATCTGTGGGACA 1011
Qy      223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
Db      1012 GGGCCAGTGTCTGAACCCAGAGACTAAAGCTGTGGCAGCTTCATAGAGACCAAGAGGAT 1071
Qy      243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 262
Db      1072 GATATTTTGTGCTTCCTGACCATGCACTCTTATGGCAGTTAATTTCTCACACCTTACGGC 1131
Qy      263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
Db      1132 TACACCAAAATAAATCAAGTAACCCAGAAATGATTCAAGTTGGACAGAGGCGACGA 1191
Qy      283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
Db      1192 AATGCATTG--AAAGCAAGATATGGAAACCAATTATAGAGTTGGATCGAGTCAGATATT 1248
Qy      303 LeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 322
Db      1249 TTATATGCCCTCATCAGGGCTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCTCTCA 1308
Qy      323 PheThr 324
Db      1309 TATACG 1314
```

Search completed: January 2, 2004, 10:05:11
Job time : 845.067 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:47:17 ; Search time 2994.02 Seconds
(without alignments)
2743.777 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FOSGQVLAALPRTSRQVQL.....IKYFTSNPPVKKLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-O=/cpn2.1/USPTO.spool/US09980881/runat_31122003_135916_26873/app_query.fasta_1.1628
-DB=EST -QFMT=fstap -SUFFIX=ret -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CGN 1 1 5167 @runat_31122003_135916_26873 -NCPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_estc2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1494	83.1	1446	11	AK004045	AK004045 Mus muscu
2	1253.5	69.7	971	9	AL536059	AL536059 AL536059
3	988	54.9	853	12	B1144644	B1144644 602909932
4	978	54.4	852	10	BG569281	BG569281 602988573
5	945.5	52.6	1004	14	BY704293	BY704293 BY704293
6	926.5	51.5	738	10	BG616456	BG616456 602614618
7	924	51.4	1022	12	B1332659	B1332659 602984408
8	921	51.2	788	10	BG618629	BG618629 602646064
9	912	50.7	747	10	BG565100	BG565100 602583720
10	903	50.2	702	2	HS089979	Bx496754 Homo sapi
11	897	49.9	747	10	BG566051	BG566051 602582652
12	894.5	49.7	724	9	AI048167	AI048167 ud71b12.y
13	863	48.0	735	9	AV646979	AV646979 AV646979
14	862	47.9	725	10	BG618239	BG618239 602645967
15	831.5	46.2	688	9	AV651709	AV651709 AV651709
16	828.5	46.1	806	10	BG567264	BG567264 602589745
17	827.5	46.0	577	14	CB154355	CB154355 K-EST0212
18	818	45.5	541	14	CB157619	CB157619 K-EST0216
19	814.5	45.3	750	10	BG618813	BG618813 602646186
20	812.5	45.2	973	9	AL536068	AL536068 AL536068
21	806.5	44.9	661	9	AI529872	AI529872 ui83dl1.y
22	806	44.8	580	9	AI182113	AI182113 ud73f10.y
23	806	44.8	589	9	AA968307	AA968307 uc71h07.y
24	795	44.2	510	14	CB156851	CB156851 K-EST0215
25	785	43.7	691	10	BG568240	BG568240 602587149
26	776.5	43.2	603	10	AW950700	AW950700 EST382770
27	766	42.6	511	9	AI574117	AI574117 u1768d09.y
28	759	42.2	511	9	AI528146	AI528146 ui94a04.y
29	755	42.0	583	9	AV693037	AV693037 AV693037
30	752	41.8	485	13	BX091627	BX091627 BX091627
31	751.5	41.8	843	10	BG216526	BG216526 RST36106
32	744	41.4	602	9	AV692032	AV692032 AV692032
33	741	41.2	601	9	AV658390	AV658390 AV658390
34	735	40.9	623	9	AV694058	AV694058 AV694058
35	732.5	40.7	537	14	N98450	N98450 za28c09.r1
36	732	40.7	434	14	CB164014	CB164014 K-EST0225
37	727.5	40.5	749	10	BG617911	BG617911 602645566
38	725	40.3	621	9	AV698425	AV698425 AV698425
39	715.5	39.8	958	9	AI322676	AI322676 m124d06.y
40	712	39.6	787	14	CB593837	CB593837 AGENCOURT
41	707	39.3	610	9	AI574064	AI574064 u167g11.y
42	698.5	38.8	723	9	AW413090	AW413090 uc52c03.x
43	688	38.3	487	9	AI574057	AI574057 u167g04.y
44	685.5	38.1	772	9	AI255929	AI255929 ui83dl1.x
45	676.5	37.6	692	10	BF384322	BF384322 602046670

ALIGNMENTS

RESULT 1
AK004045
LOCUS
DEFINITION
AK004045 1446 bp mRNA linear HTC 05-DEC-2002
Mus musculus 18-day embryo whole body cDNA, RIKEN full-length
enriched library, clone:110032P04 product:carboxypeptidase B2
(plasma), full insert sequence.
ACCESSION
AK004045 GI:12835067
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
HTC; CAP trapper.
Mus musculus (house mouse)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;


```
QY 61 HisLeuAenValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLeu 80
Db 285 CATTTAAATGTGAGCAAGATTCCATTAACTGATGAACACGTGGAGGACCTTAAT 344
QY 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 345 GAACAGCAGACTTTCATGACCGGTGAGCCCGCGCCCTCCGCTTCATACATAGAGCAG 404
QY 101 TyrHisSerLeuAenGluIleTyrSerTyrPheGluPheIleThrGluArgHisProAsp 120
Db 405 TATCACTCGCTAAATGAATCTATCTCTGATAGAAGTCACTAAGTGAACAGCATCTCGAC 464
QY 121 MetLeuThrIleHisIleGlySerSerPheGluIleTyrProLeuTyrValLeuIle 140
Db 465 ATGCTCCAGAAATCTACATCGGATCATCATCTCGAAGATACCACCTTATATGTTTAAAG 524
QY 141 ValSerGlyIleGluGlnThrAlaIleAsnAlaIleTyrPheIleAspCysGlyIleHisAla 160
Db 525 GTCTCAGAAAGGACAAAGAAATCAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 584
QY 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis 175
Db 585 AGAGAATGGATTTCACCTGCTTCTGTGTGTGTCTATAGGCTACGTGACACAAATCCAT 644
QY 175 ----- 175
Db 645 GGGAAAGAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 704
QY 176 ----- 183
Db 705 ATGAACGTGGATGGCTATGACTACACGTGGAAGAAATCGATGTGGAGGAAGACCGC 764
QY 184 SerPheTyrAlaAenHisCysIleGlyThrAspLeuAenArgAsnPheAlaSerLys 203
Db 765 TCTGCTCACAAAGAACACCGCTCGTGGGCACAGACCTGAACAGGAATTCGCTTCCAA 824
QY 204 HisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db 825 CACTGTGTGAGAAAGGTGGTCAAGTTCCTCTGCTCTGAAACCTACTGTGGACTTAT 884
QY 224 ProGluSerGluProGluValIleAlaValAlaSerPheLeuArgArgAsnIleAenGln 243
Db 885 CTTGAGTCTGAGCCAGAGGTGAAGGACGTGGCTGACTTCTTGAGAAAGAAATATCGACCAC 944
QY 244 IleIleAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
Db 945 ATTAAGCTTACATCATGATGACATCATCTCCACAAATACTGTTCCCTATTCCTAT 1004
QY 264 ThrArgSerLysSerLysAspHisGluGluSerLeuValAlaSerGluAlaValArg 283
Db 1005 AACAGAAGCAAAAGCAAGGACCAAGAACTGTCTCTAGTGGCCAGCGAGCAGTTCGT 1064
QY 284 AlaIleGluIleThrSerLysAenThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db 1065 GCAATTGAAGATTAATAAAACACACGAGTACACACGCGAGTGGCTCAGAAAGTTA 1124
QY 304 TyrLeuAlaProGlyGlyIleAspAspTyrIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db 1125 TATCTAGCTCTGGAGGTTCTGAGATTTGATTTGGCATCAATATTCGTT 1184
QY 324 -----ThrSerAs 326
Db 1185 ACAATTGAGCTCCGAGATACAGGCAGATACGGATTTCTGTGCTGTGAGATACATAA 1244
QY 326 nProProValGluIleLeuLeuProLeuSerLeuIle 338
Db 1245 CCCACTGTGCAGAAAGCTTTGGCGGCCATCTCTAAA 1281
RESULT 2
LOCUS AL536069 971 bp mRNA linear EST 31-MAY-2003
DEFINITION AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF022YE21 5-PRIME, mRNA sequence.
```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL536069
AL536069.2 GI:31260939
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 971)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12799562.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6944.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF022AC11QP1&cluster=6944.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0DF022AC11QP1.
Location/Qualifiers
1. 971
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 282 a 201 c 211 g 272 t 5 others
ORIGIN

FEATURES
Source

Alignment Scores:
Pred. No.: 6.72e-132 Length: 971
Score: 1253.50 Matches: 240
Percent Similarity: 86.02% Conservative: 0
Best Local Similarity: 86.02% Mismatches: 2
Query Match: 69.72% Indels: 37
DB: 9 Gaps: 1
US-09-980-881A-3 (1-338) x AL536069 (1-971)
QY 1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 131 TTTCAAGATGGCCCAAGTTCTAGCTGCTTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 190
QY 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuLeu 40
Db 191 CAGACTTCTACTACAACATATGAGATTGTTCTCTGGCAGCGGTAAACAGCTGACCTTATT 250
QY 41 ValLysIleGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60
Db 251 GTCAAGAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCCGCAATGTGAAAGCC 310
QY 61 HisLeuAenValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuLeu 80
Db 311 CATTTAAATGTGAGCGGAATTCATGAGTGTCTTCTGGCAGACGTGGAGATCTTATT 370
QY 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db 371 CAACAGCAGATTTCACACGACACAGTCAGCCCGAGCCCTCCGCATCGTACTATGAACAG 430
QY 101 TyrHisSerLeuAenGluIleTyrSerTyrPheGluPheIleThrGluArgHisProAsp 120

Db 431 TATCACTCACTAAATGAAATCTATCTCTGGATAGATTTTATAACTCAGAGGCATCTCTGAT 490

Qy 121 MetLeuThrIleHisIleGlySerSerPheGluIysTyrProLeuTyrValLeuLys 140

Db 491 ATGCTTACAAAATCCACATGGATCTCTATTTGAGAAAGTACCACACTCTATGTTTAAAG 550

Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrPheLeuAspCysGlyIleHisAla 160

Db 551 GTTCTGGAAGAAGACAGCAGCAAAATGCCATATGGATTGACTGTGGAAATCCATGCC 610

Qy 161 ArgGluTyrPheLeuSerProAlaPheCysLeuTyrPheIleGlyHis 175

Db 611 AGAGAAATGGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670

Qy 175 ----- 175

Db 671 GGGATAATAGGGCAATATACCAATCTCTCAGGCTTGTGGATTTCTATGTTATGCCGGTG 730

Qy 176 -----AsnArgMetTyrArgLysAsnArg 183

Db 731 GTTAATGTGATGTTATGACTACTCATGGAAGAAAGATCGAATGTGGAGAAAGACCGT 790

Qy 184 SerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203

Db 791 TCTTTCTATCGAACAATCATTCATTGCAATGGAACAGACCTGAATAGGAACCTTTGCTCCAA 850

Qy 204 HisTyrCysGluGluGlyAlaSerSerSerSerSerSerSerSerSerSerSerSerSerSer 223

Db 851 CACTGGTGTGAGGAAGTGATCCAGTTCCTCATGCTCGGAACCTACTGTGCACTTTAT 910

Qy 224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242

Db 911 CCTGAGTCAGAACAGAGTGAAGGACGTGCTAGTCTTGTGAGAGAAATTAAC 967

RESULT 3
BI144644
LOCUS 602909932F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5051358 5',
DEFINITION mRNA sequence.
ACCESSION BI144644
VERSION BI144644.1 GI:14604645
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 853)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1139 row: O column: 07
High quality sequence stop: 849.
Location/Qualifiers
1. 853
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5051358"
/lab_host="DHIOB (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
/notes="Organ: liver; Vector: pCMV-SF06T6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life

FEATURES

source

Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 243 a 199 c 194 g 217 t

ORIGIN

Alignment Scores:

Pred. No.: 9,96e-102 Length: 853
Score: 988.00 Matches: 194
Percent Similarity: 78.28% Conservative: 15
Best Local Similarity: 72.66% Mismatches: 18
Query Match: 54.95% Indels: 40
DB: 12 Gaps: 2

US-09-980-881A-3 (1-338) x BI144644 (1-853)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Db 59 TTTCAAGATGGCCAGGTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 118

Qy 21 GlnAsnLeuThrThrTyrGluIleValLeuTyrGlnProValThrAlaAspLeuIle 40

Db 119 CAGATCTTACTACACGATATGAGTCTCTCTGCGCCAGTCAGAGCTGATTCATC 178

Qy 41 ValLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60

Db 179 GAGAAGAAAAGGAAGTCCACTTTTGTGAATGCGTCTGATCTGCACAGTGTCAAAGCG 238

Qy 61 HisLeuAsnValSerGlyIleProCysSerValLeuAlaAspValGluAspLeuIle 80

Db 239 CATTTAAATGTGACGAGAATTCATTTAACTGTAACCAACCTGGAGAGCACTAATT 298

Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Db 299 GAACACAGACATTTCAATGACAGCTGACGCCCGCGCCCTCCGCTTCATCTATGAGCAG 358

Qy 101 TyrHisSerLeuAsnGluIleTyrSerTyrPheIleGluPheIleThrGluArgHisProAsp 120

Db 359 TATCACTCGCTAAATGAAATCTATTCTCGATAGAAGTCATACTGAACAGCATCTCTGAC 418

Qy 121 MetLeuThrIleHisIleGlySerSerPheGluIysTyrProLeuTyrValLeuLys 140

Db 419 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCACCTTTATGTTTAAAG 478

Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrPheIleAspCysGlyIleHisAla 160

Db 479 GTCTCAGGAAGGAACAAAGATCAAAATGCCATCTGGATCGACTGTGGAAATCCATGCC 538

Qy 161 ArgGluTyrPheLeuSerProAlaPheCysLeuTyrPheIleGlyHis 175

Db 539 AGAGAATGGATTTCACTGCTTTCTGTTGTGTTCTATAGGCTACGTGACACAAATCCAT 598

Qy 175 ----- 175

Db 599 GGGAAAAGAAATCTGTATACAGACTTCTGAGGCAGCTGGATTTCTACATCATGCCCGTG 658

Qy 176 -----AsnArgMetTyrArgLysAsnArg 183

Db 659 ATGAACGTGATGCTGATGACTACAGCTGGAAAAGAAATCGAATGTGGAGAGAACCGT 718

Qy 184 SerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203

Db 719 ---CTGTCAAGAAACAAACCGTTCGTTGGGCCAGACCTGAACAGGAACCTTCGCTTCCAA 775

Qy 204 HisTyrCysGluGluGlyAlaSerSerSerSerSerSerSerSerSerSerSerSerSerSer 223

Db 776 CACTGGTGTGAGAAAGGTGCTCCAGTTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 835

Qy 223 YrProGluSerGluPro 228

Db 836 ATCTGAGTCTGAGCCG 852

RESULT 4
BG569281
LOCUS BG569281 852 bp mRNA linear EST 10-APR-2001

DEFINITION	602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5', mRNA sequence.
ACCESSION	BG569281
VERSION	BG569281.1 GI:13576934
KEYWORDS	EST.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens (human)
REFERENCE	
AUTHORS	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	NIH-MGC http://imgc.nci.nih.gov/.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCMI580 row: b column: 19 High quality sequence stop: 751.
FEATURES	Location/Qualifiers
source	1..852 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4722354" /lab_host="DH10B (T1 phage-resistant)" /clone_lib="NIH_MGC_76" /notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1: SfiI (ggccattatggc); Site 2: SfiI (ggccattatggc); 5' and 3' adaptors are used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT	251 a 177 c 187 g 237 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1.37e-100 Length: 852
Score:	978.00 Matches: 197
Percent Similarity:	76.15% Conservative: 1
Best Local Similarity:	75.77% Mismatches: 5
Query Match:	54.39% Indels: 58
DB:	10 Gaps: 2
US-09-980-881A-3 (1-338) x BG569281 (1-852)	
Qy	135 ProLeuTyrrValLeuTyrsValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIle 154
Db	2 CCACCTATGTTTAAAGGTTTCTGGAAGAAGCAAGCAGCCAAATGCGATGATGAT 61
Qy	155 AspCysGlyLleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGly 174
Db	62 GACTGTGAATCCATGCCAGAGAATGATCTCTCTGCTTCTGCTTCTGCTTCATAGGC 121
Qy	175 His----- 175
Db	122 CATATACTCAATCTATGGGATAATAGGCAATATACCAATCTCTGAGCGTTGGAT 181
Qy	176 -----AsnArg 177
Db	182 TTCTATGTTATGCCAGTGGTTATATGCGATGGTTATGACTACTCATCGGAAGAATCGA 241
Qy	178 MetTrp-ArgLysAsnArgSerPheTyrrAlaAsnHisCysIleGlyThrAspLeuAs 197
Db	242 ATGTGACGAAAGAACCGTTCTTCTATCGAACATCATTCATCGGAACACGACCTGAA 301
Qy	197 nArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerG 217
Db	302 TAGGAACCTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCTCTCATGCTCGGA 361
Qy	217 uThrTyrrCysGlyLeuTyrrProGluSerGluProGluValLysAlaValAlaSerPheLe 237
Db	362 AACCTACTGTGACTTTTATCTCTGAGTCAGAACAGGAAGTGAAGGAGTGGCTAGTTCTT 421
Qy	237 uArgArgAsnIleAsnGlnIleLysAlaTyrrIleSerMetHisSerTyrrSerGlnHisil 257
Db	422 GAGAAGAAATATCAACAGCATTAAGCATACATCAGCATGCATTCATCTCCAGCATAT 481
Qy	257 eValPheProTyrrSerTyrrThrArgSerLysAspHisGluGluLeuSerLeuVa 277
Db	482 AGTGTTCCTATATTCCTATACAGAGTAAGTAAGCAAGCAAGCAAGCAAGCAAGCAAG 541
Qy	277 lalaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrrThrHisG 297
Db	542 AGCCAGTGAAGCAGTTTCGTGCTATTGAGAAAATTAGTAAATAATACAGGTATACACATGC 601
Qy	297 yHisGlySerGluThrLeuTyrrLeuAlaProGlyGlyGlyAspAspTrpIleTyrrAspL 317
Db	602 GCATGCTCAGAAACCTTATACCTAGTCTCTGAGGTGGGAGCAGTGGATCTATGCATT 661
Qy	317 euGlyIleLysTyrrSerPhe----- 323
Db	662 TGGGCATCAGATATTCTGTTTACAATTTGAACTTCGAGATACGGGCACATAACGATTCTT 721
Qy	324 -----ThrSerAsnProProValGluLysLeuProLeuSerLeu 337
Db	722 GCTGCCGGAGCGTTACATCAAA-CCACCTGTAGAGAAGCTTTTGGCGGTCTCTCTA 776
RESULT 5	
BY704293	
LOCUS	1004 bp mRNA linear EST 16-DEC-2002
DEFINITION	BY704293 RIKEN full-length enriched, 18-day embryo whole body Mus
ACCESSION	musculus cDNA clone 1110032P04 5', mRNA sequence.
VERSION	BY704293.1 GI:27115417
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 1004) Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusci, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragan, T. A., Fletcher, C. F., Forrest, A., Frazer, J. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Qi, D., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynnshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12468851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda
, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,
Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno
, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,
Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
Submission

**Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences** Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

**Computer-based methods for the mouse full-length cDNA
encyclopedia:** real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

Location/Qualifiers
1. 1004
/organism="Mus musculus"
/mol type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1110032P04"
/tissue type="whole body"
/dev stage="18-day embryo"
/clone lib="RIKEN full-length enriched, 18-day embryo
whole body"
286 a 228 c 228 g 262 t

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 8,95e-97 Length: 1004
Score: 945.50 Matches: 188
Percent Similarity: 76.69% Conservative: 16
Best Local Similarity: 70.68% Mismatches: 24
Query Match: 52.59% Indels: 41
DB: 14 Gaps: 2

US-09-980-881A-3 (1-338) x BY704293 (1-1004)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
105 TTTCCAGAGTGCCAGGTTTATCTGCTTCTCCAGAACCTCCAGGCAAGTTCAACTACTT 164
Db 21 GlnAsnLeuThrThrThrGluLeuValLeuThrGlnProValThrAlaAspLeu 40
165 CAGAAATCTTACTCAACAGTATAGGTCGTTCTCTGGCAGCAGCAGTCAATTCATC 224
Qy 41 ValLysLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60

Db 225 GAGAGAAAAAGGAAGTCCACTTTTGTGAATGGCTCTGTGATCTGCACAGTGTCAAAGCG 284
Qy 61 HisLeuAsnValSerGlyLeuProCysSerValLeuAlaAspValGluAspLeu 80
285 CATTTAAATGTGAGCAGAAATCCATTTAAAGTTCGATGAACAAGTGGAGACCTTAAT 344
Qy 81 GlnGlnGlnLeuSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
345 GAAACAGCAGACTTCAATGACAGCGTCAGCCCCCGCGCTTCGCTTCACTATGAGCAG 404
Db 101 TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
405 TATCACTCGCTAAATGAAATCTATTCTCGATAGAGTCAATACTGAACAGCATCTCGAC 464
Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
465 ATGCTCCAGAAATCTACATCGGATCATCATTCGAGAAAGTACCCACTTTATGTTTAAAG 524
Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
525 GTCCTAGGAAGGACAAAGAAATCAAAATGCCATCTGGATCGACTGTGAATCCATGCC 584
Qy 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
585 AGAGAATGGAATTCACCTGCTTCTGTTGTGTTCTATAGGCTACGTCACACAATTCAT 644
Qy 175 ----- 175
Db 645 GGGAAAGAAAAATCTGTATACACAGACTTCTGAGCAGCTGGATTTTACATCATCGCGTG 704
Qy 176 ----- AsnArgMetTrpArgLysAsnArg 183
705 ATGAACGTGGATGGCTATGACTACAGCTGGAAAAAGAAATCGTATGTGGAGAGAGCGC 764
Qy 184 SerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
765 TCTGCTCACAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACCTCGCTTCCAA 824
Qy 204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
825 CACTGGTGTGAAG---GTGCGTAAGTTCTCTGCTCTGA-ACCTACTGTGGAAT-TAT 879
Qy 224 ProGluSerGluProGlu 229
Db 880 TCTAG-TCTGAGCCAGAG 896

RESULT 6
LOCUS BG616456 738 bp mRNA linear EST 18-APR-2001
DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
mRNA sequence.
ACCESSION BG616456
VERSION BG616456.1 GI:13667827
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 738)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LLCM1595 row: c column: 05

```

FEATURES
source
High quality sequence stop: 666.
Location/Qualifiers
1. 738
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:473500"
/lab_host="DH10B (TI phage-resistant)"
/clone_lib="NIH_MGC_76"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
sfii (ggcgctcgcc); Site 2: sfii (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTAGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCGGCAGATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

BASE COUNT 209 a 161 c 153 g 215 t
ORIGIN
Alignment Scores:
Pred. No.: 8.12e-95 Length: 738
Score: 926.50 Matches: 192
Percent Similarity: 71.13% Conservatives: 10
Best Local Similarity: 67.61% Mismatches: 27
Query Match: 51.53% Indels: 55
DB: 10 Gaps: 3

US-09-980-881A-3 (1-338) x BG616456 (1-738)
QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 40 TTTCAGATGCGCAAGTTCTAGCTGCTTCTTAGAACCTCTAGGCAAGTTCAAGTTCTA 99
QY 21 GlnAsnLeuThrThrThrTyrrGluLeuValLeuTrpGlnProValThrAlaAspLeuLe 40
Db 100 CAGAATCTTACTCAACATATGAGATTGTCTCTGGCAGCGGTAACAGCTGACCTTATT 159
QY 41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspValLysAla 60
Db 160 GTGAAGAAAAACAGTCCCAATTTTGTAAATGCACTCTGATGTCACAAATGGAAGCC 219
QY 61 HisLeuAsnValSerGlyLeuProCysSerValLeuAlaAspValGluAspLeuLe 80
Db 220 CATTTAAATGTAGCGGAATTCATGACAGTCTTCTGCGACAGCTGGAAGATCTTATT 279
QY 81 GlnGlnGlnSerAsnAspThrValSerProArgAlaSerAlaSerTyrrTyrrGluGln 100
Db 280 CAACAGCAGATTTTCCAAACGACACAGTCCAGCCCGAGCCCTCCGATCGTACTATGAACAG 339
QY 101 TyrHisSerLeuAsnGluLeuTyrrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db 340 TATCACTCACTAATGAATCTATTCTTGTGATAGAAATTTTATACTGAGAGGCATCTGTAT 399
QY 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrrProLeuTyrrValLeuLys 140
Db 400 ATGCTTACAAAATCCACATGATGATCTCTATTGAGAGTACCACCTCATGTTTAAAG 459
QY 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db 460 GTTCTTGGAAAGAACAAAGCAGCAAAATATGCCATATGGATTGACTGTGGAATCCATGCC 519
QY 161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--AsnArgMetTrpA 180
Db 520 AGAGAATGGATCTCTCTGCTTTCTGCTGTTTCTATAGGCCATATAACTCAATCTTAT 579
QY 180 rGlyAsnArgSerPheThrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnP 200
Db 580 GGGATATAGGCAATATACCAATCTCTCTGAGCGCTGTGGA----- 620
QY 200 heAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrrC 220

```

```

Db 621 -----TTTCTATGTTATGCCGGTGTCA-----T 645
QY 220 ySGlyLeuTyrrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAga 240
Db 646 GTGGACTG----- 653
QY 240 snIleAsnGlnIleLysAlaTyrrIleSerMetHisSerTyrrSerGlnHisIleValPheP 260
Db 654 -----GTTATGG 660
QY 260 roTyrrSerTyrrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
Db 661 ACTACTCATGGACCAAGATCGCATTGTTGGACAAAAGAACCCGGTCTTTACTAATGCG 720
QY 280 luAlaVal 282
Db 721 AAACAATT 728

RESULT 7
BI332659 1022 bp mRNA linear EST 30-JUL-2001
LOCUS 602984408F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5137369 5',
DEFINITION mRNA sequence.
ACCESSION BI332659
VERSION BI332659.1 GI:15017316
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://img.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11337 row: 0 column: 02
High quality sequence stop: 841.
Location/Qualifiers
1. 1022
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5137369"
/lab_host="DH10B (TI phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 303 a 233 c 225 g 261 t
ORIGIN
Alignment Scores:
Pred. No.: 2.59e-94 Length: 1022
Score: 924.00 Matches: 200
Percent Similarity: 73.67% Conservatives: 21
Best Local Similarity: 66.67% Mismatches: 37
Query Match: 51.39% Indels: 45
DB: 12 Gaps: 2

US-09-980-881A-3 (1-338) x BI332659 (1-1022)
QY 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20

```



```

Db      552 GTTCTGGAAGAACAGCAGCCAAATGGCATATGGATTGACTGTGGAATCCATGCC 611
Qy      161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis-----175
Db      612 AGAAGATGGATCTCTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 671
Qy      175 -----175
Db      672 GGGATAATAGGCAATATTACCAATCTCTCTGAGGCTGTGGAATTTCTATGTCAGT 731
Qy      176 -----AsnArgMetTrpArgLys 181
Db      732 GGTAAATGTTGGATGGTTATGACTACTCATGGAAGAAAGCAATGATGTGGAGAA 788

RESULT 9
LOCUS   BG565100              747 bp mRNA linear EST 10-APR-2001
DEFINITION BG565100.1 GI:13572753
ACCESSION BG565100
VERSION   BG565100.1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NTH-MGC http://mgi.nci.nih.gov/
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLCMI551 row: h column: 05
High quality sequence stop: 714.
Location/Qualifiers
1. 747
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4711348"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: liver; Vector: pDNR-LJB (Clontech); Site 1:
SfiI (ggcgccctcgcc); Site 2: SfiI (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-ATTCTAGAGCCGAGCGGCATG-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGGCATG-3' (30 BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."
BASE COUNT 220 a 154 c 166 g 207 t
ORIGIN

Alignment Scores:
Pred. No.: 3,72e-93 Length: 747
Score: 912.00 Matches: 196
Percent Similarity: 80.48% Conservative: 6
Best Local Similarity: 78.09% Mismatches: 7
Query Match: 50.72% Indels: 45
DB: 10 Gaps: 2

US-09-980-881A-3 (1-338) x BG565100 (1-747)

```

```

Qy      30 ValLeuTrpGlnProValThrAlaAspLeuIleValLysLysLysGlnValHisPhePhe 49
Db      1 GTTCTCTGCAGCCGGTAACAGCTGACCTTAT-GTGAAGAAAAAACAAGTCCATTTT 59
Qy      50 ValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCys 69
Db      60 GTRGATGCAATCTGATGTCACAAATGTGAAGCCCAATTAATGTGAGCGAATTCATGC 119
Qy      70 SerValLeuLeuAlaAspValGluAspLeuIleGlnGlnIleSerAsnAspThrVal 89
Db      120 AGTGCTCTTCTGTCAGATGTGAAGATCTTATTTCAACAGCAGATTTCCACAGCAGTC 179
Qy      90 SerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSer 109
Db      180 AGCCCCGAGCCCTCCGATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCT 239
Qy      110 TrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySer 129
Db      240 TGGATAGAATTTATAACTGAGAGGCATCTTGATATGCTTACAAAAATCCCATTTGATCC 299
Qy      130 SerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLys 149
Db      300 TCATTTGAGAAGTACCCTCATCTATGTTTAAAGGTTTCTGAAAAAGCAAGCAGCCAAA 359
Qy      150 AsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCys 169
Db      360 AATGCCATATGGATTGACTGTGAATCCATGCCAGAAATGGACCTCTCTCTGCTTTCTGC 419
Qy      170 LeuTrpPheIleGlyHis-----175
Db      420 TTGTGTTTCATAGGCCATATACTCAATTCATTTGGGATAATAGGCAATATACCAATCTC 479
Qy      176 -----AsnArgMet 178
Db      480 CTGAGGCTTGTGGATTCTTATGTTATGCCAGTGTTAATGTGGATGGTATGACTACTCA 539
Qy      179 TrpArg-----LysAsnArgSerPheTyrAlaAsnHisCys 191
Db      540 TGGAAAAAGAAATCGGAATGTGGAGAACAGAACCGTCTTCTTATGCGAACATCATTTGC 599
Qy      192 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGlu-GlyAlaSe 211
Db      600 ATCGGAACAGACCTGAATAGGAATT-GCTTCCAAACACTGTGTGAGGAAGGTCATC 658
Qy      211 rSerSerSer-CysSerGlu-ThrTyr-CysGlyLeuTyrProGluSerGluProGluVa 230
Db      659 CAGTCTCTCATGCTCGAAACCTACTTGTGGACTTTATCTGTGAGTTCAGAACAGAGTG 718
Qy      230 lLysAlaValAlaSerPheLeuArg 238
Db      719 -GAAGCAGTGGCTAGTTCTTGTGAAG 742

RESULT 10
HSM089979
ID      HSM089979 standard; RNA; EST; 702 BP.
XX      AC BX496754;
XX      AC BX496754;
XX      SV BX496754.1
XX      09-MAY-2003 (Rel. 75, Created)
DT      09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX      Homo sapiens mRNA; EST DKFZp779I1529_r1 (from clone DKFZp779I1529)
DE      EST; expressed sequence tag.
XX      Homo sapiens (human)
OS      Homo sapiens (human)
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX      Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX      [1]
RN      1-702
RP

```

RA	Ansonger W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA	Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RT	Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL	MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
RL	
XX	This is the 5' sequence of the clone insert
CC	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC	Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC	sequenced by EMBL (European Molecular Biology Laboratories,
CC	Heidelberg/Germany) within the cDNA sequencing consortium of
CC	the German Genome Project.
CC	No sl sequence available.
CC	This clone (DKFZp779i1529) is available at the RZPD in Berlin.
CC	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC	14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX	
XX	Key Location/Qualifiers
PH	
FT	source
FT	1. .702
FT	/db xref="taxon:9606"
FT	/mol_type="mRNA"
FT	/organism="Homo sapiens"
FT	/clones="DKFZp779i1529"
FT	/clone_lib="779 (synonym: hnccl). Vector pSport1_sfi; host:
FT	DH10B; sites SfiIIA + SfiIB"
FT	/dev stage="fetal"
FT	/dev stage="fetal"
FT	/issue_type="liver"
XX	
SQ	Sequence 702 BP; 197 A; 149 C; 146 G; 210 T; 0 other;
Alignment Scores:	
Pred. No.:	3.59e-92 Length: 702
Score:	903.00 Matches: 174
Percent Similarity:	98.86% Conservative: 0
Best Local Similarity:	98.86% Mismatches: 1
Query Match:	50.22% Indels: 1
DB:	2 Gaps: 0
US-09-980-881A-3 (1-338) x HSM089979 (1-702)	
Qy	1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db	125 TTTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 184
Qy	21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
Db	185 CAGAATCTTACTACAAATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 244
Qy	41 ValLys-LysLysGlnValHisPhePheValAsnAlaSerAspValAsnValLysAl 60
Db	245 GTGAAGAAATAAACAAGTCCATTTTTTTGTAATGCACTGATGTCGACAAATGGAAGC 304
Qy	60 aHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuI 80
Db	305 CCATTTAAATGTGAGCGGAATTCATGCAGTGTCTTGCTGGCAGCCGTGGAAGATCTTAT 364
Qy	80 eGlnGlnGlnIleSerIleAspThrValSerProArgAlaSerAlaSerTyrTyrGluGl 100
Db	365 TCAACAGCAGATTTTCCAAACGACACAGTCAGCCCGCGAGCTCCGCACTGCTACTGAACA 424
Qy	100 nTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAs 120
Db	425 GTATCACTCACTAAATGAATCTATTTCTGGATAGAAATTTATAACTGAGAGCATCTCGA 484
Qy	120 pMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLy 140
Db	485 TATGCTTACAAAATCCACATTTGATCTCTCATTTGAGAGTACCCACTCATGTTTTTAAA 544
Qy	140 sValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAl 160
Db	545 GGTTTCTCGAAAGAACAAAGCAGGCCAAAATGCCATATGGATTGCACTGTGGAATCCATGC 604

Qy	160	aArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
Db	605	CAGAGAATGATCTCTCCGTTCTGCTTGTGTTATAGGCAT 650
RESULT 11		
BG566051		
LOCUS		
DEFINITION		602582652F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5', mRNA linear EST 10-APR-2001
ACCESSION		BG566051
VERSION		1
KEYWORDS		GI:13573704
SOURCE		Homo sapiens (human)
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS		1 (bases 1 to 747)
TITLE		NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT		Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM1549 row: g column: 13 High quality sequence stop: 731.
FEATURES		Location/Qualifiers
source		1..747 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:4710564" /lab_host="DH10B (T1 phage-resistant)" /clone_lib="NIH MGC 76" /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgctgggcc); Site_2: SfiI (ggcattaggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 clones contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."
BASE COUNT		212 a 157 c 159 g 219 t
ORIGIN		
Alignment Scores:		
Pred. No.:		1,91e-91 Length: 747
Score:		897.00 Matches: 178
Percent Similarity:		92.35% Conservative: 3
Best Local Similarity:		90.82% Mismatches: 12
Query Match:		49.89% Indels: 3
DB:		10 Gaps: 0
US-09-980-881A-3 (1-338) x BG566051 (1-747)		
Qy	1	PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db	86	TTTCAGAGTGGCCAAAGTTCTAGCTGCTCTCTCTATTAACCTCTAGGCAAGTTCAAGTTCTA 145
Qy	21	GlnAsnLeuThrThrThrTyrrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
Db	146	CAGAACTTACTACACATATGAGATTGTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205
Qy	41	ValLysLysLysGlnValHisPhePheValAsnAlaSerValAspValAsnValysAla 60
Db	206	GTCAAGAAAAAACAAGTCAATTTTTTTTAAATGTCATCTGATCTGCATGATGTGAAGCC 265

Db 662 TCTGCTCACAGAACCACCGCTGCTGGGCACAGACCTGAACAGGAATTCGCTCCAAA 721

QY 204 His 204
|||

Db 722 CAC 724

RESULT 13
AV646979

LOCUS AV646979 735 bp mRNA linear EST 15-JAN-2002

DEFINITION AV646979 GLC Homo sapiens cDNA clone GLCATD06 3', mRNA sequence.

ACCESSION AV646979

VERSION AV646979.1 GI:9867993

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 735)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z. and Han, Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

PUBMED 11752456

COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..735
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCATD06"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 200 a 158 c 172 g 205 t

ORIGIN

Alignment Scores:
Pred. No.: 1.39e-87 Length: 735
Score: 863.00 Matches: 170
Percent Similarity: 96.59% Conservative: 0
Best Local Similarity: 96.59% Mismatches: 5
Query Match: 48.00% Indels: 1
DB: 9 Gaps: 0

US-09-980-881A-3 (1-338) x AV646979 (1-735)

QY 1 PhcInSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Db 58 TTTTCAGATGCGCAAGTTCTAGTGTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTA 117

QY 21 GlnAsnLeuThrThrThrThrGluLeuValLeuTrioGlnProValThrAlaAspLeu 40

Db 118 CAGAACTTACTACACATATGAGATTGTTCTTGGCAGCCGTAACAGGTGACCTTATT 177

QY 41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspValLysAla 60

Db 178 GTCAAGAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTCGACATGTGAAGCC 237

QY 61 HisLeuAsnValSerGlyLeuProCysSerValLeuLeuAlaAspValGluAspLeu 80
|||||

Db 238 CATTTAAATGTGAGCGGAATTCATGCAGTGTCTGCTGGCAGACGTGAAGATCTATT 297

QY 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
|||||

Db 298 CAACAGCAGATTTCCAACGACACAGTCAGCCGCCCGCATCCCGCATCGGACATATGAACAG 357

QY 101 TyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
|||||

Db 358 TATCACTACTAAATGAAATCTATTCTGGATAGAATTTTATACTGAGAGGCATCTCGAT 417

QY 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
|||||

Db 418 ATGCTTACAAAATCCACATTCCTCATTTTGAGAGTACCACCTCTATGTTTAAAG 477

QY 141 -ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyLeHisAl 160
|||||

Db 478 GGTGTCTCGAAAGAACAAACAGCCAAATAATGCCATATGATGACTGTGGAATCCATGC 537

QY 160 aArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
|||||

Db 538 CAGAGAAATGATCTCTCTGCTTTCTGTGGTGGGGGTCTATAGGCCAT 583

RESULT 14

LOCUS BG618239

DEFINITION BG618239 725 bp mRNA linear EST 18-APR-2001

ACCESSION BG618239

VERSION BG618239.1 GI:13669610

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 725)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1629 row: f column: 17
High quality sequence stop: 723.

FEATURES
source
1..725
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4767376"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctggcc); Site 2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCCATATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGCCGACATG-(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."

BASE COUNT 203 a 153 c 149 g 220 t

ORIGIN

Alignment Scores:

Pred. No.: 1-77e-87 Length: 725
 Score: 862.00 Matches: 172
 Percent Similarity: 98.29% Conservative: 0
 Best Local Similarity: 98.29% Mismatches: 3
 Query Match: 47.94% Indels: 2
 DB: 10 Gaps: 0

US-09-980-881A-3 (1-338) x BG518239 (1-725)

Qy 1 PheGlnSerGlyGlnValLeuAlaLeuProArgThrSerArgGlnValGlnValLeu 20
 Db 131 TTTTCAGAGTGGCCCAAGTCTTAGCTCTTCTAGAACCTCTAGGCAAGTTCAAGTTCTTA 190
 Qy 21 GlnAsnLeuThrThrThrTyrGluLeuValLeuTyrGlnProValThrAlaLeuLeu 40
 Db 191 CAGAATCTTACTACACATATGAGAT-GTTCTCTGGCAGCGGTAAACAGCTGACCTTAT- 248
 Qy 41 ValLeuLysLysGlnValHisPhePheValAsnAlaSerValAspAsnValLysAla 60
 Db 249 GTGAAGAAACCAAGTCCATTTTTTTTAAATGCACTGATGTCGACAAATGGAAGCC 308
 Qy 61 HisLeuAsnValSerGlyLeuProCysSerValLeuAlaAspValGluAspLeuLeu 80
 Db 309 CATTAAATGTGAGCGGAATTCATGAGTCTTCTGTCGAGATGTGGAAGATCTTATT 368
 Qy 81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
 Db 369 CAACAGCAGATTTCCAAACGACACAGTCAGCCCCGAGCGCTCCGCACTGCTATGAACAG 428
 Qy 101 TyrHisSerLeuAsnGluLeuTyrSerTyrPheLeuPheLeuThrGluArgHisProAsp 120
 Db 429 TATCACTCACTAAATCAATCTATTCTTGATAGAAATTTATACTGAGAGGCACTCTGAT 488
 Qy 121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
 Db 489 ATGCTTACAAATCCACATGGATCTCTCATTTGAGAAGTACCCTCTATGTTTAAAG 548
 Qy 141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrPheLeuAspCysGlyLeuHisAla 160
 Db 549 GTTCTCGAAAGAAACAGCAGCAAAATGCCATATGGATTGACTGTGGAATCCATGCC 608
 Qy 161 ArgGluTyrIleSerProAlaPheCysLeuTyrPheIleGlyHis 175
 Db 609 AGAGAATGGATCTCTCTGCTTCTGTTGTTTCATAGGCCAT 653

RESULT 15

AV651709 688 bp mRNA linear EST 15-JAN-2002
 AV651709 GLC Homo sapiens cDNA clone GLCCSF10 3', mRNA sequence.

AV651709
 VERSION AV651709.1 GI:9872723

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 688)
 Xue, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
 Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W.,
 Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
 Hu, G., Gu, J., Chen, Z. and Han, Z.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

MEDLINE 21625106

PUBMED 11752456

COMMENT Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.

FEATURES

source Location/Qualifiers
 1..688
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCCSF10"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 208 a 145 c 151 g 184 t
 ORIGIN

Alignment Scores:

Pred. No.: 4-89e-84 Length: 688
 Score: 831.50 Matches: 162
 Percent Similarity: 89.50% Conservative: 0
 Best Local Similarity: 89.50% Mismatches: 1
 Query Match: 46.25% Indels: 18
 DB: 9 Gaps: 1

US-09-980-881A-3 (1-338) x AV651709 (1-688)

Qy 176 AsnArgMetTyrArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 195
 Db 46 ATCTGATCTGGAGAAAGAACCGTCTTCTATGCGAACATCATTCGATCGAACAGAC 105
 Qy 196 LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGluAlaSerSerSerCys 215
 Db 106 CTGAATAGGAACCTTGTCTTCCAAACACTGGTGTGAGAAAGTGCATCCAGTTCCTCATGC 165
 Qy 216 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaAlaSer 235
 Db 166 TCGGAACCTACTCTGTGAGACTTTATCTCTGAGTCAAGAACCAAGTGAAGGAGTGGCTAGT 225
 Qy 236 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 255
 Db 226 TTCTTGAGAGAAATATCAACAGATTAAAGCATACATCAGCATGCATTCATATCCAG 285
 Qy 256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 275
 Db 286 CATATAGTGTTCATATTCCTATATACCAAGTAAAGCAAGCAACCATGAGGAAGTGTCT 345
 Qy 276 LeuValAlaSerGluAlaValAlaGluAlaIleGluLysThrSerLysAsnThrArgTyrThr 295
 Db 346 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAACTAGTAAAAATACCAGGTATACA 405
 Qy 296 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTyrPheTyr 315
 Db 406 CATGGCCATGGCTCAGAAACCTTTATACCTTAGCTCTCTGGAGGTGGGACGATGATCTAT 465
 Qy 316 AspLeuGlyIleLysTyrSerPhe----- 323
 Db 466 GATTTGGCATCAAAATATTCGTTTACAATTGAACCTCGAGATACCGGCACATACCGATTC 525
 Qy 324 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLeu 338
 Db 526 TTTTGTGGCGAGCGGTGTACATCAAAACCCACCTGTTGTGAGAGCTTTTGGCGCTGTCTCTAA 585
 Qy 338 s 338
 Db 586 A 586

Search completed: January 2, 2004; 02:37:30

Job time : 3005.02 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:33:22 ; Search time 268.597 Seconds
(without alignments)
2472.334 Million cell updates/sec

Title: US-09-980-881a-4
Perfect score: 1338
Sequence: 1 ASASYEQHSLNEIYSWIE.....IKYFTSPNPVKLLPLSLK 246

Scoring table:

BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Ygapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/QFMT-fasta -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-DB=N_geneseq_19Jun03 -QFMT=fasta -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=PTO -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=PTO -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881@cgn 1.1 577 @runat 31122003 135914 26819 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1325	99.0	1573	21	AAC81962	Human brain carbox
2	1289.5	96.4	1625	25	ABX13670	Human protease CDN
3	1272	95.1	1272	20	AAV74302	Human plasma carbo
4	1272	95.1	1728	24	ABN95899	Gene #2397 used to
5	1272	95.1	1749	17	AAT11671	Human plasma carbo
6	1272	95.1	1749	18	AAT62846	Human plasma carbo
7	1267	94.7	1749	14	AAQ41001	Human plasma carbo
8	902	67.4	1400	21	AAF18005	Lung cancer associ
9	610	45.6	927	17	AAT35760	Rat mature carboxy
10	582	43.5	921	16	AAQ90601	Porcine carboxypep
11	582	43.5	1215	16	AAQ90600	Porcine Tyr-His-Me
12	567	42.4	999	17	AAT42500	Mature HCPB coding
13	567	42.4	1053	17	AAT42497	Mature HCPB (His)6
14	567	42.4	1263	17	AAT42494	Human pancreatic c
15	567	42.4	1263	19	AAV41795	Human pancreatic c
16	567	42.4	1284	17	AAT42506	ProHCPB gene with
17	560	41.9	1053	20	AAZ24804	Human carboxypepta
18	560	41.9	1059	17	AAT42511	Modified HCPB (D25
19	560	41.9	1059	18	AAT62787	Carboxypeptidase B
20	559	41.8	1059	17	AAT42512	Modified HCPB (D25
21	559	41.8	1059	18	AAT62788	Carboxypeptidase B
22	556.5	41.6	1302	25	ABV77254	Nucleotide sequenc
23	556.5	41.6	1311	22	AAD03837	Human carboxypepti
24	556.5	41.6	1993	18	AAD30579	Human protease, PR
25	556	41.6	1059	18	AAT62806	Carboxypeptidase B
26	554	41.4	1059	18	AAT62790	Carboxypeptidase B
27	554	41.4	1059	18	AAT62801	Carboxypeptidase B
28	553	41.3	1059	18	AAT62789	Carboxypeptidase B
29	552	41.3	1059	18	AAT62791	Carboxypeptidase B
30	552	41.3	2023	25	ABZ75116	Anti-human seminal
31	551	41.2	1059	18	AAT62805	Carboxypeptidase B
32	551	41.2	1059	18	AAT62803	Carboxypeptidase B
33	551	41.2	1059	18	AAT62804	Carboxypeptidase B
34	550.5	41.1	1125	22	AAT62807	Human secreted met
35	550.5	41.1	1125	22	AAT76477	CDNA encoding huma
36	550.5	41.1	1125	24	ABK31745	DNA encoding novel
37	550.5	41.1	1222	25	ABT33353	NOVX DNA sequence
38	550.5	41.1	1342	22	AAH76476	Human secreted met
39	550.5	41.1	1342	22	AAH76478	Human secreted met
40	550.5	41.1	1344	25	ABQ77393	Human zinc-binding
41	550.5	41.1	1378	25	ABT33355	NOVX DNA sequence
42	550.5	41.1	1603	24	AAD39053	Human 23566 (carbo
43	550.5	41.1	1826	22	AAF81728	Human protease and
44	549	41.0	1059	18	AAT62808	Carboxypeptidase B
45	549	41.0	1059	18	AAT62800	Carboxypeptidase B

ALIGNMENTS

RESULT 1

AAC81962
ID AAC81962 standard; cDNA; 1573 BP.

XX AAC81962;

XX 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B cDNA.

XX Carboxypeptidase B; human; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX Homo sapiens.
OS

DR WPI: 2003-102387/09.
DR P-PSDB; ABG72093, ABG72095.
XX New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
PT
PT
PT
XX
PS
XX
XX
XX
CC The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for
CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorders. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC cDNA, the gene for which is located on chromosome 13.
XX
SQ Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;

Alignment Scores:
Pred. No.: 6.15e-138 Length: 1625
Score: 1289.50 Matches: 244
Percent Similarity: 92.42% Conservative: 0
Best Local Similarity: 92.42% Mismatches: 2
Query Match: 96.38% Indels: 18
DB: 25 Gaps: 1

US-09-980-881A-4 (1-246) x ABX13670 (1-1625)

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 20
DB 359 GCCTCCGCATCGTACTATGACAGATATCACTCACTAAATGAAATCTATTCTGGATAGAA 418
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 419 TTTATTAAGTGAAGGCATCTGATATGCTTACAAAATCCATTTGGATCTCTATTGAG 478
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 479 AAGTACCACCTCTATGTTTAAAGTTTCTGGAAGAAGCAACAGCAGCAAAATGCCATA 538
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
DB 539 TGGATTGACTGTGAATCCATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGGTTC 598
QY 81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
DB 599 ATAGGCCATTAATCGAATGTGGAGAAGAAGACCGTCTTCTATCGCAACATCATTTGATC 658
QY 101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
DB 659 GGAACAGACCTGATAGAACTTTGCTTCCAAACACTGGTGTGAGGAGGTGATCCAGT 718
QY 121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140

DB 719 TCCTCATGCTCGGAAACCTACTGTGACTTTATCTCTGAGTCAGAACCAAGAGTGAAGGCA 778
QY 141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
DB 779 GTGGCTAGTTTCTTGAGAAAGAAATATCAACACAGATTAAAGCATATCATCAGCATGCAATCA 838
QY 161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
DB 839 TACTCCAGCATATAGTGTTCATATTCTATACAGAAAGTAAAGCAAGACCATGAG 898
QY 181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
DB 899 GAACCTGCTCTAGTACGAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAATATACC 958
QY 201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220
DB 959 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACGAT 1018
QY 221 TrpIleTyrAspLeuGlyIleLysTyrSerPhe----- 231
DB 1019 TGGATCTATGATTGGGCATCAAAATATTCTTTTACAAATTGAACTTCGAGATACGGGCACA 1078
QY 232 -----ThrSerAsnProValGluLysLeuLeuProLe 243
DB 1079 TACGGATTCTTGTCTGCCGAGCGTTACATCAAAACCCACTGTAGAGAGCTTTTGGCGCT 1138
QY 243 uSerLeuLys 246
DB 1139 GTCTCTAAAA 1148
RESULT 3
AAV74302
ID AAV74302 standard; cDNA; 1272 BP.
XX
AC AAV74302;
XX
DT 28-APR-1999 (first entry)
DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
DE
KW Plasma carboxypeptidase B: PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease; ds.
XX
OS Homo sapiens.
XX
PN WO9855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
XX WPI: 1999-045800/04.
DR P-PSDB; AAW92270.
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Example 1; Page 24; 35pp; English.
XX
CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2

CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.

XX
SQ Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

Alignment Scores:

Pred. No.: 4,41e-136 Length: 1272
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 20 Gaps: 2

US-09-980-881A-4 (1-246) x AAV74302 (1-1272)

Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 20
Db 343 GCTCCGACATCGTACTATGACAGTATCACTCAATAATGAATCTATTCTTGATAGAA 402
Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 403 TTTATACTGAGAGGCATCTGATATGCTTACAAAATCCACATTTGGATCTCATTTGAG 462
Qy 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 463 AAGTACCACCTCTATGTTTAAAGGTTTCTGGAAGAAACAAACAGCCAAATATGCCATA 522
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 523 TGGATTGACTGTGAATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTTTGGTTC 582
Qy 81 IleGlyHis----- 83
Db 583 ATAGGCCATATACTCAATCTATGGGATAATAGGCAATATACCAATCTCTGAGGCTT 642
Qy 83 ----- 83
Db 643 GTGGATTCTATGTTATGTCGGTGGTTAATGTGACGGTTATGACTACTCATCGAAGAAAG 702
Qy 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 703 AATCGAATGTGGAAGAAAGACCGTCTTCTATGCGAACAATCATTTGTCATCGGAACAGAC 762
Qy 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 763 CTGATAGGAACCTTGTCTTCAACACTGTGTGAGGAGGTGCATCCAGTTCTCATGC 822
Qy 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 823 TCGGAACCTTACTGTGAGCTTTATCTCGTAGTCAAGACCAGAGTGAAGCGAGTGTAGT 882
Qy 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 883 TTCTTGAGAAGAATAATCAACAGATTAAAGCATACATCAGCATGATCATCATCTCCAG 942
Qy 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 943 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAACCATGAGGAACCTGTCT 1002
Qy 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1003 CTAGTAGCCAGTAGGACAGTTCTGTCTATTGAGAAACCTAGTAAATAATACAGATACA 1062
Qy 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyr 223
Db 1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGGTGGGACGATGATCTAT 1122
Qy 224 AspLeuGlyIleLysTyrSerPhe----- 231

Db 1123 GATTGGGCATCAAAATATTGTTTACAAATTCGAGATACGGGCACATACGGATTC 1182
Qy 232 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeu 246
Db 1183 TTGCTGCCGAGCGTTACATCAACCCACCTGTAGAGAAAGCTTTTGGCGGTCTCTATAA 1242
Qy 246 s 246
Db 1243 A 1243
RESULT 4
ABN95899
ID ABN95899 standard; DNA; 1728 BP.
XX
AC ABN95899;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2397 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
FN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
DR
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
XX liver tissue sample -
PS Claim 1; SEQ ID NO 2397; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;
XX
Alignment Scores:
Pred. No.: 6,93e-136 Length: 1728
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 24 Gaps: 2

```
US-09-980-881A-4 (1-246) x ABN95899 (1-1728)
OY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 20
Db 362 GCCTCCGCATCGTACTATGAAACAGTATCACTCAATAATGAAATCTATTCTCGATAGAA 421
OY 21 PheIleThrGluArgHisProAspMetLeuThrIleHisIleGlySerSerPheGlu 40
Db 422 TTTATAACTGAGAGGATCCGTATATGCTTACAAAAATCCACATTCGATCCTCATTTGAG 481
OY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 482 AAGTACCCACTCTATGTTTAAAGTTTCTGAAAAGAACAAACAGCCAAATATGCATA 541
OY 61 TrpIleAspCysGlyIleHisAlaArgLysTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 542 TGGATTGACTGTGGAATCCATGCCAGAGATGATCTCTCTGCTTTCTGCTTGTGGTTC 601
OY 81 IleGlyHis----- 83
Db 602 ATAGGCCATATACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTT 661
OY 83 ----- 83
Db 662 GTGGATTCTATGTTATGCGGTGTTAATGTGGACGGTTATGACTACTCATCGAAAAAG 721
OY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 722 AATCGAATGTGGAGAAAGAACCGTCTTCTATGCGACAAATCATTCGATCGGAACAGAC 781
OY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 782 CTCAATAGGAACCTTGTCTTCAACACATGCTGTGAGGAGGTGCATCCAGTTCCTCATGC 841
OY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 842 TCGGAACCTTACTGTGGACTTTATCTGTAGTCAGAACAGAACTGAAGCGAGTGGCTAGT 901
OY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 902 TTCCTTGAGAGAAATATCAACACAGATTAAAGCATACATCAGCATGCATTCATCTCCAG 961
OY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 962 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGAACACCATGAGGAACGTCT 1021
OY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAAGTATACA 1081
OY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTyrTrpIleTyr 223
Db 1082 CATGGCATGCTCAGAAACCTTATACCTAGCTCTCTGGAGTGGGACGATGATCTAT 1141
OY 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 1142 GATTGGGCATCAAAATATTGTTTACAAATTGAACTTCGAGATACGGGCACATACGGATTC 1201
OY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
Db 1202 TTGCTGCGGAGGCTTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAA 1261
OY 246 s 246
Db 1262 A 1262
RESULT 5
AAT11671
XX ID AAT11671 standard; DNA; 1749 BP.
AC AAT11671;
XX XX
```

```
DT 25-MAR-2003 (updated)
DE 12-APR-1996 (first entry)
XX Human plasma carboxypeptidase B coding sequence.
XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
CDS 41..1312
FT /*tag= a
FT /product= Human plasma carboxypeptidase B.
FT sig_peptide 41..106
FT mat_peptide /*tag= b
FT 107..1309
FT /*tag= c
XX US5474901-A.
PN 12-DEC-1995.
XX 19-JUL-1994; 94US-0277540.
XX 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
XX (GETH ) GENENTECH INC.
XX Drayna DT, Eaton DL;
PI WPI: 1996-0319508/04.
DR P-PSDB; AAR90293.
XX Antibody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A
XX Disclosure; Figure 4; 40pp; English.
XX An antibody which specifically binds human plasma carboxypeptidase B
CC (hPCPB) and does not cross react with other carboxypeptidases is
CC useful for the detection of hPCPB in vitro. The antibody is also
CC used for purifying hPCPB from a sample. Purification comprises
CC passing a sample thought to contain hPCPB over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hPCPB.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX SQ Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Alignment Scores:
Pred. No.: 7,05e-136 Length: 1749
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 55
Query Match: 95.07% Indels: 2
DB: 17 Gaps: 2

US-09-980-881A-4 (1-246) x AAT11671 (1-1749)
OY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 20
Db 383 GCCTCCGCATCGTACTATGAAACAGTATCACTCAATAATGAAATCTATTCTCGATAGAA 442
OY 21 PheIleThrGluArgHisProAspMetLeuThrIleHisIleGlySerSerPheGlu 40
Db 443 TTTATAACTGAGAGGATCCGTATATGCTTACAAAAATCCACATTCGATCCTCATTTGAG 502
```


Qy	41	LysTyrProLeuTyrValLeuLysValSerGlyLysGluInThrAlaLysAsnAlaIle	60
Db	503	AAGTACCACCTCTATGTITTTAAAGGTTCTGGAAAAGAACAAACAGCAAAAAATGCCATA	562
Qy	61	TripLeaspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe	80
Db	563	TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCGTGCTTCTGCTTGTGGTTC	622
Qy	81	IleGlyHis-----	83
Db	623	ATAGCCCATATAAECTCAATTCTATGGGTAATAGGGCAATATACCATCTCTCTGAGGCTT	682
Qy	83	-----	83
Db	683	GTGGATTCTATGTTATGTCGGTGCGTGTAAATGTCGGACGGTTATGACTACTCATCGAAAAAG	742
Qy	84	AsnArgMetTrpArgLysAsnArgSerPheTyrZlalaSnAsnHisCysIleGlyThrAsp	103
Db	743	AATCGAATGTGGAGAAAGAACCGTTCCTTCTATCGGAACAATCATTCGATCGGAACAGAC	802
Qy	104	LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerCys	123
Db	803	CTGAATAGGAACCTTTGGTCTCAAAACACTGGTGTGAGGAAGTGCAATCCAGTTCTCTATGC	862
Qy	124	SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer	143
Db	863	TCGAAACCTTACTCTGGACTTTATCTCTGAGTCAGAACCCAGAAAGTGAAGGCAGTGGCTAGT	922
Qy	144	PheLeuArgArgAsnIleAsnGlnIleIysAlaTyrlleserMetHisSerTyrSerGln	163
Db	923	TTCTTGAGAAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCAATTCATACTCCCAG	982
Qy	164	HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer	183
Db	983	CATATAGTGTTTCCATATTCCATACACGAAGTAGTAAAGCAAAGCAATGAGGAACCTGCT	1042
Qy	184	LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr	203
Db	1043	CTAGTACCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAAGTATACA	1102
Qy	204	HisGlyHisGlySerGluThrLeuTyrIleuAlaProGlyGlyIleAspAspTrpIleTyr	223
Db	1103	CATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCGAGGTGGGGACGATTGGATCTAT	1162
Qy	224	AspLeuGlyIleLysTyrSerPhe-----	231
Db	1163	GATTGGGCATCAAAATATTGTTTACAACTTCGACATTCGACATACGGGCACATACGATTC	1222
Qy	232	-----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys	246
Db	1223	TTGCTGCGGAGCGTTACATCAAAACCCACCTGTAGAGAAGCTTTGCGGCTGCTCTCTAAA	1282
Qy	246	s 246	
Db	1283	A 1283	
 RESULT 6 AAT62846			
ID	AAT62846 standard; DNA; 1749 BP.		
XX	AAT62846;		
XX			
DT	25-MAR-2003 (updated)		
DT	08-MAY-1997 (first entry)		
XX			
DE	Human plasma carboxypeptidase B coding sequence.		
XX			
KW	Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;		
KW	plasma; plasminogen, 58.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	

```

CDS
FT 41..1312
FT /*tag= a
FT /product= Human PCPB
FT 41..106
FT /*tag= b
FT mat_peptide
FT 107..1309
FT /*tag= c
FT misc_binding
FT 134..177
FT /*tag= d
FT /bound_moeity= 46_bp_probe
XX
PN US5593674-A.
XX
XX 14-JAN-1997.
XX
XX 27-APR-1995; 95US-0430787.
XX
XX 01-FEB-1991; 91US-0649591.
XX 14-OCT-1992; 92US-0959944.
XX 15-DEC-1993; 93US-0167727.
XX 19-JUL-1994; 94US-0277540.
XX 27-APR-1995; 95US-0430787.
XX
XX (GETH ) GENENTECH INC.
XX
XX Drayna DT, Eaton DL;
XX
XX WPI; 1997-099413/09.
XX P-PSDB; AAW14733.
XX
XX Using human plasma carboxy:peptidase B in blood coagulation - is
XX functionally related to carboxy:peptidase A and pancreas
XX carboxy:peptidase B
XX
XX Example 2; Column 37-42; 39pp; English.
XX
XX This sequence encodes human plasma carboxypeptidase B (PCPB) which
XX has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
XX PCPB may be used therapeutically in haemostatic regulation. PCPB is
XX purified from human plasma or by transformed cell culture by
XX extraction using plasminogen bound to a solid phase.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
XX SQ Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Alignment Scores:
Pred. No.: 7.05e-136 Length: 1749
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 18 Gaps: 2

US-09-980-881A-4 (1-246) x AATG2846 (1-1749)
Qy 1 AlaserAlaserTVrTVrGluGlnTVrHisSerLeuAsnGluIleTVrSerrTrpIleGlu 20
Db 383 GCCTCCGCATCGTACTATGAACAGATACACTACTAAATGAATCTATCTTGGATAGAA 442
Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 443 TTTATAACTCAGAGGCATCTCGATATGCTTACAAAAATCCACATTTGGATCTCTATTGG 502
Qy 41 LysTVrProLeuTVrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 503 AAGTACCACCTCTATGTGTATTAAGGGTTCTCGAANAAGAACAAACAGCCAAAAATGCCATA 562
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 563 TGGATTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGGTTC 622
Qy 81 IleglyHis----- 83

```

```
Db 623 ATAGCCATATACTCAATCTTATGGGATAATAGGCAATATACCAATCTCCTGAGGCTT 682
QY 83 ----- 83
Db 683 GTGGATTCTATGTTATGCGGTGTTAATGTGCGGTTTATGACTACTCATGAAAAAG 742
QY 84 AsnArgMetTrrArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 743 AATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACCAATCATTCATCGGAACAGAC 802
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrrCysGluGluGlyAlaSerSerSerCys 123
Db 803 CTGAATAGGAACCTTGCTTCCAAACACTGGTGTGTGAGGAAGTGCATCCAGTCTCTCATGC 862
QY 124 SerGluThrTrrCysGlyLeuTrrProGluSerGluProGluValLysAlaValAlaSer 143
Db 863 TCGGAACCTTACTGTGAGCTTTATCTGAGTTCAGAACCAAGAGTGAAGCGAGTGTAGT 922
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTrrLysSerMetHisSerTrrSerGln 163
Db 923 TTCTTGAGAAAGAAATATCAACAGATTAAAGCATATACATCAGCATGCATTCATCTCCAG 982
QY 164 HisIleValPheProTrrSerTrrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 983 CATATAGTGTTCATATCTTATACACGAAGTAAAGCAAGCAACCATGAGGAACTGTCT 1042
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTrrThr 203
Db 1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAACTAGTAAATAATACAGGTATACA 1102
QY 204 HisGlyHisGlySerGluThrLeuTrrLeuAlaProGlyGlyGlyAspTrrIleTrr 223
Db 1103 CATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACGATGGATCTAT 1162
QY 224 AspLeuGlyIleLysTrrSerPhe----- 231
Db 1163 GATTGGGCATCAATATTCGTTTCAATTAAGTTCGAGATACGGGCACATACGATTC 1222
QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
Db 1223 TTGCTGCGGAGGCTTATACATCAAAACCACCTGTAGAGAGCTTTGCGGCTGTCTATA 1282
QY 246 s 246
Db 1283 A 1283
RESULT 7
ID AAQ41001 standard; cDNA; 1749 BP.
XX AAQ41001;
AC AAQ41001;
XX XX
DT 24-AUG-1993 (first entry)
XX XX
DE Human plasma carboxypeptidase B gene.
XX XX
KW PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX XX
OS Homo sapiens.
XX XX
FH Key Location/Qualifiers
FT sig_peptide 41..106
FT FT /*tag= a
FT mat_peptide 107..1312
FT FT /*tag= b
FT FT /product= PCPB
FT FT 134..177
FT FT /*tag= c
FT FT /function= probe
FT FT /note= "used to obtain full-length clones"
XX XX
PN US5206161-A.
XX XX
```

```
PD 27-APR-1993.
XX PF 01-FEB-1991; 91US-0649591.
XX PR 01-FEB-1991; 91US-0649591.
XX PA (GETH ) GENENTECH INC.
XX PI Drayna DT, Eaton DL;
XX WIPI; 1993-151724/18.
XX DR P-PSDB; AAR36273.
XX PT New human plasma carboxypeptidase B - used as haemostatic
XX PT regulator for clotting blood, partic. for treating blood clotting
XX PT disorders, e.g. haemophilia
XX PS Disclosure; Fig 4; 40pp; English.
XX CC Human plasma carboxypeptidase B was isolated from human plasma and
XX CC partially sequenced. Oligonucleotide primers were designed based on
XX CC the partial amino acid sequences. The primers were used in a PCR
XX CC amplification to identify cDNA encoding PCPB from a human liver cDNA
XX CC library. The PCR product was capable of encoding the first 37 amino
XX CC acids of PCPB; a 46mer probe was used to obtain the full-length
XX CC sequence which, although disclosed in the specification, is not
XX CC claimed. PCPB inhibits the enzymatic conversion by tPA of
XX CC plasminogen to plasmin in the presence of fibrinogen.
XX SQ Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

Alignment Scores:
Pred. No.: 2.65e-135 Length: 1749
Score: 1267.00 Matches: 245
Percent Similarity: 81.40% Conservative: 0
Best Local Similarity: 81.40% Mismatches: 1
Query Match: 94.69% Indels: 55
DB: 14 Gaps: 2

US-09-980-881A-4 (1-246) x AAQ41001 (1-1749)
QY 1 AlaSerAlaSerTrrTrrGluGlnTrrHisSerLeuAsnGluIleTrrSerTrrIleGlu 20
Db 383 GCCTCCGCATCGTACTATGACAGATATCACTCAATAAGAAATCTATTCTGGATAGA 442
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 443 TTTATAACTGAGGCGCATCTGTATATGCTTACAAAATCCACATTTGGATCTCTATTGAG 502
QY 41 LysTrrProLeuTrrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 503 AAGTACCCACTCTATGTTTAAAGGTTTCTGGAAAAGAACAAACAGCAAAATGCCATA 562
QY 61 TrrIleAspCysGlyIleHisAlaArgGluTrrIleSerProAlaPheCysLeuTrrPhe 80
Db 563 TGGATTGACTGTGGAATCCATGCGCAGAAATGGATCTCTCCTGCTTTCTGCTTGGTTC 622
QY 81 IleGlyHis----- 83
Db 623 ATAGGCCATATACTCAATTTCTATGGGCAATATAGGCAATATACCAATCTCCTGAGGCTT 682
QY 83 ----- 83
Db 683 GTGGATTCTATGTTATGCGGTGTTAATGTGCGGTTTATGACTACTCATGAAAAAG 742
QY 84 AsnArgMetTrrArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 743 AATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACCAATCATTCATCGGAACAGAC 802
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrrCysGluGluGlyAlaSerSerSerCys 123
Db 803 CTGAATAGGAACCTTGCTTCCAAACACTGGTGTGTGAGGAAGTGCATGAGTCTCTCATGC 862
```

QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
 DB 863 TCGGAACCTACTGTGACCTTATCTGAGTCAGAACCCAGAGTGAAGCAGTGGCTAGT 922
 QY 144 PheLeuArgArgAenlleAenGlnlleLysAlaTyrIleSerMetHisSerTyrSerGln 163
 DB 923 TTCTTGAGAAAGAAATATCAACAGATTAAAGCATATACATCAGCATGCATTCTATCTCCAG 982
 QY 164 HisIleValPheProTyrSerTyrThrArgSerIysSerIysAspHisGluLeuSer 183
 DB 983 CATATAGTGTTCATATCTTACACAGAGTAAAGCAAGCAACCATGAGGAAGTGTCT 1042
 QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAenThrArgTyrThr 203
 DB 1043 CTAGTAGCCAGTAGAGCAGTTCGTCTATTGAGAAACTAGTAAATATACAGGTATACA 1102
 QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTyrIleTyr 223
 DB 1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGACGATTGGATCTAT 1162
 QY 224 AspleuGlyIleLysTyrSerPhe----- 231
 DB 1163 GATTTGGGCATCAAAATATTCGTTTACAAATTGAATTCGAGATACGGGCACATACGGATTC 1222
 QY 232 -----ThrSerAenProProValGluLysLeuLeuProLeuSerLeuLys 246
 DB 1223 TTGCTGCGGAGCGTTACATCAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTTAA 1282
 QY 246 s 246
 DB 1283 A 1283
 RESULT 8
 ID AAF18005 standard; DNA; 1400 BP.
 XX AC AAF18005;
 XX AC AAF18005;
 DT 14-MAR-2001 (first entry)
 XX DE Lung cancer associated polynucleotide sequence SEQ ID 24.
 XX KW Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerable;
 KW gastrointestinal; nephrotropic; antineoplastic; gynecological;
 KW antibacterial; diagnostic; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX OS Homo sapiens.
 XX PN W0200055180-A2.
 XX PD 21-SEP-2000.
 XX PF 08-MAR-2000; 2000WO-US05918.
 XX PR 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX (ROSE/) ROSEN C A.
 XX PI Ruben SM;
 XX WPI; 2000-587514/55.
 XX P-PSDB; AAB58129.
 XX Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX Claim 1; Page 507; 1425pp; English.
 XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer

CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active; general; vulnerable; gastrointestinal
 CC general; nephrotropic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer.
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterization of the polynucleotide and protein
 CC sequences.
 XX

SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;

Alignment Scores:

Pred. No.: 1.8e-93 Length: 1400
 Score: 902.00 Matches: 181
 Percent Similarity: 76.05% Conservative: 0
 Best Local Similarity: 76.05% Mismatches: 1
 Query Match: 67.41% Indels: 56
 DB: 21 Gaps: 2

US-09-980-881A-4 (1-246) x AAF18005 (1-1400)

QY 65 GlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTyrPheIleGlyHis--- 83
 DB 10 GGAATCATGCCAGAAATATGGAATCTCTCTGCTTTCTGCTTGGTTTCATAGGCCATATA 69
 QY 83 ----- 83
 DB 70 ACTCAATTTCTATGGGATAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTAT 129
 QY 84 -----AsnArgMetTyr 87
 DB 130 GTTATGCCGTGGTTAATGTGGATGTTATGAATCTACTATGGAAGAAAGATCGAATGTG 189
 QY 87 PARGLysAsnArgSerPheTyrAlaAenAenHisCysIleGlyThrAspLeuAenArgAs 107
 DB 190 GAGAAAGAACCGTCTCTTTCTATGCGAACATATTCATCGAACAGACAGCTTGAATAGAA 249
 QY 107 nPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerSerCysSerGluThrTy 127
 DB 250 CTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAACCTA 309
 QY 127 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 147
 DB 310 CTGTGGACTTTATCTGAGTCAGAACCCAGAACCCAGAACCCAGAACCCAGAACCCAGAAC 369
 QY 147 gAenlleAenGlnlleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 167
 DB 370 AAATATCAACAGATTAAAGCATATACATCAGCATGCATTCTATCTCCAGCATATAGTGT 429
 QY 167 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 187
 DB 430 TCCATATTTCTATACAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 489
 QY 187 rGluAlaValArgAlaIleGluLysThrSerLysAenThrArgTyrThrHisGlyHisGl 207
 DB 490 TGAAGCAGATTCTGCTATTGAGAAACCTAGTAAATAATACAGGATATACATGAGCCATGG 549
 QY 207 YSerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyI 227
 DB 550 CTGAGAAACCTTATACCTAGCTCCTGGAGTGGGACGATTGGATCTATGATTGGCAT 609
 QY 227 eLysTyrSerPhe----- 231
 DB 610 CAAATATTCGTTTACAATTCGAGATACGGGCACATACGAGTTCCTGTGTCGCGGA 669

QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
 Db 670 GCCTTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGCTGTCTCTAAAA 719

RESULT 9

AAT35760
 ID AAT35760 standard; cDNA; 927 BP.

XX AAT35760;
 XX 09-NOV-1996 (first entry)
 XX Rat mature carboxypeptidase B cDNA.
 XX Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.
 XX Rattus sp.

PH Key Location/Qualifiers
 FT primer_bind /tag= a
 FT /note= "Mature CPB 5'-end primer"
 FT primer_bind 898..927
 FT /tag= b
 FT /note= "CPB 3' end primer"

XX W09623064-A1.

PN 01-AUG-1996.

XX 25-JAN-1996; 96WO-US000995.

XX 25-JAN-1995; 95US-0378233.

XX (BIOT-) BIOTECHNOLOGY GEN CORP.

PI Fulga N, Gorecki M, Hartman J, Mendelovitch S;

XX WPI; 1996-362688/36.

DR P-PSDB; AAW00602.

XX Purified active recombinant carboxypeptidase B prodn. - by
 PT expressing DNA encoding the pro-enzyme, folding and enzymatic
 PT cleavage to give active form, used e.g. for insulin prodn.

XX Example 1; Page 37-38; 49pp; English.

XX A cDNA sequence (AAT35760) codes for the mature form (AAW00602) of
 CC rat carboxypeptidase B (CPB). A full-length sequence including
 CC cDNA (AAT35759) coding for the activation peptide (AAW00601) can
 CC be obt'd. by PCR amplification (see also AAT35756 and AAT35758) of cDNA
 CC derived from Sprague-Dawley rat pancreas. Pro-CPB can be expressed
 CC in host cells, e.g., Escherichia coli, and subsequently recovered,
 CC refolded and cleaved with trypsin to yield the active enzyme. CPB
 CC produced this way is cheaper than porcine pancreatic enzyme, and is
 CC free of other proteases.

XX Sequence 927 BP; 253 A; 226 C; 213 G; 235 T; 0 other;

Alignment Scores:

Pred. No.: 3-72e-60 Length: 927
 Score: 610.00 Matches: 126
 Percent Similarity: 60.50% Conservative: 44
 Best Local Similarity: 44.84% Mismatches: 71
 Query Match: 45.59% Indels: 41
 DB: 17 Gaps: 5

US-09-980-881A-4 (1-246) x AAT35760 (1-927)

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpLeuGlu 20

Db 1 GCAAGTGGACACAGCTACACCAAGTACAACTGGGAAACGATTGAGCGGTGATTCAA 60

QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
 Db 61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCAATGGAAACCAATTGAA 120
 QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
 Db 121 GGACGTAACATGATGTCTCTCAAGATT---GGTAAACTAGACCGGAATAGCCTGCCATC 177
 QY 61 TrpIleAspCysGlyIleHisAlaArgGluTTrpIleSerProAlaPheCysLeuTrpPhe 80
 Db 178 TTCAATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCCTCTGCAATTCGTGAGTGGTTT 237
 QY 81 Ile----- 81
 Db 238 GTGAGAGAGGCTGTCCGTACTTATATCAAGAGATCCACATGAACAGCTTCTAGATGAA 297
 QY 82 -----GlyHis----- 83
 Db 298 CTGGATTCTATGTTCTGCTGTGGTCAACATTGATGGCTATGCTACACCTGGACTAAG 357
 QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
 Db 358 GACAGAATGTGGAGAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTGGGTGTAGAC 417
 QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
 Db 418 CCCACAGGAATTTT---AATGCTGCTGTGTGAAGTGGAGCTTCTCGAGTCCCTGC 474
 QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
 Db 475 TCTGAACCTTACTGTGGACAGCCCGAGAGTCTGAAAAGAGACAAGGCCCTGGCAGAT 534
 QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMethIleSerTyrSerGln 163
 Db 535 TTCAATCGCAACCACTCTCCACCATCAAGGCTACTGACCATCCACTCATCTACTACAG 594
 QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
 Db 595 ATGATGCTCTACCTTACTCTATGACTACAACTGCCTGAGAACTATGAGGAATGAAT 654
 QY 184 LeuValAlaSerGluAlaValAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
 Db 655 GCCTGTGTAAGGTGCGGAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711
 QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
 Db 712 TATGGCCCGAGGAGCTACAACTATCTATCTGCTGGGGGATCTGACGACTGGTCTTAT 771
 QY 224 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 243
 Db 772 GATCAGGAATCAAAATATTCTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830
 QY 244 Ser 244
 Db 831 TCT 833
 RESULT 10
 AAO90601
 ID AAO90601 standard; DNA; 921 BP.
 XX AAO90601;
 XX 13-MAR-1996 (first entry)
 DT Porcine carboxypeptidase B coding sequence.
 XX Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
 KW human serum albumin; premating factor alpha; mating factor alpha;
 KW proCBB; ds.
 XX Sus scrofa.

PH Key Location/Qualifiers
 FT 1..921
 FT /*tag= a
 FT /product= Porcine carboxypeptidase B.
 XX
 PN WO9514096-A1.
 XX
 XX 26-MAY-1995.
 XX 16-NOV-1994; 94WO-US13142.
 XX 16-NOV-1993; 93US-0153258.
 XX (ELIL) LILLY & CO ELI.
 XX Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
 PI Sterner JL, Zhang H;
 XX WPI; 1995-200386/26.
 DR P-PSDB; AAR75132.
 XX
 XX DNA encoding porcine carboxypeptidase B - used for transforming
 PT host cells, partic. Pichia species, for prodn. of the enzyme
 XX
 PS Claim 1; Page 20; 34pp; English.
 XX
 CC The porcine carboxypeptidase B coding sequence can be place in a
 CC bacterial or pref. Pichia yeast expression vector. The expression
 CC vector further comprises the signal peptide of either human serum
 CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
 CC (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
 CC pFJ474 - NRRL B-21032) or the porcine proCB signal peptide,
 CC (designated pLGD27 - NRRL B-21027). The method can be used for
 CC producing large amounts of porcine carboxypeptidase B and when
 CC produced in Pichia yeast, the protein does not need solubilisation
 CC or folding. The produced enzyme is then used for pref. cleaving
 CC basic residues from the carboxy terminus of proteins.
 XX
 SQ Sequence 921 BP; 270 A; 238 C; 194 G; 219 T; 0 other;

Alignment Scores:
 Pred. No.: 6.12e-57 Length: 921
 Score: 582.00 Matches: 117
 Percent Similarity: 59.47% Conservative: 40
 Best Local Similarity: 44.32% Mismatches: 67
 Query Match: 43.50% Indels: 40
 DB: 16 Gaps: 5

US-09-980-881A-4 (1-246) x AAQ90601 (1-921)

QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
 DB 16 TATGAGAAGTACAACTACCTGGGAAACGACGAGCTTGGACTAAGCAAGTACCAGTGAA 75
 QY 26 HisProaspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
 DB 76 AATCCAGCTCACTCTCCGACAGCCATCGGAACATACATTTTAGAACAATATATAC 135
 QY 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
 DB 136 CTCCTCAAGTT--GGCAACCTGGACCAATAAAGCTGCCATTTTCATGGACTGTGT 192
 QY 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 81
 DB 193 TTCATGCCAGAAATGGATTTCATGCAATTTTCCAGTGTGTGTGTGTGTGTGTGTGT 252
 QY 81 ----- 81
 DB 253 CTCACCTATGATATGAGTCAATGACAGAAATTCCTCAACAGCTAGACTTTATGTC 312
 QY 82 -----GlyHis-----AsnArgMetTrpArg 88
 DB 313 TTGCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGCAACCAATGTGAGA 372

QY 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
 DB 373 AAGACCGGCTTACCAATGCTGGAATCTACTGCTGGACACACCCCAACAGAAATTTT 432
 QY 109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 128
 DB 433 ---GATGCTGGGTGGTGCAACACTGGAGGCTCTACAGACCCCTGCGATGAGACTTACTGT 489
 QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
 DB 490 GGATCTGCTGCAGAGTCTGAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAAC 549
 QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
 DB 550 CTCTCTCATCAAGCATACCTGACGATCCATCATCTACATCAGATGATCTTACCT 609
 QY 169 TyrSerTyrThrArgSerLysAspHisGluGluSerLeuValAlaSerGlu 188
 DB 610 TATTCCTATGATTACAACTCCCGAGAACAAATGCTGAGTTGAATAACCTGGCTAAGCT 669
 QY 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
 DB 670 GCGGTGAAAGAACTT---GCTACACTGTATGGCAACCAAGTACATACGCGCCAGGAGCT 726
 QY 209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
 DB 727 ACAACATCTATCTCTGCTGGGGCTCTGATGACTGGCTTATGACCAAGAAATCAAA 786
 QY 229 TyrSerPheThr 232
 DB 787 TATTCCTTCCACC 798
 RESULT 11
 AAQ90600
 ID AAQ90600 standard; DNA; 1215 BP.
 XX
 AC AAQ90600;
 XX
 DT 13-MAR-1996 (first entry)
 XX
 DE Porcine Tyr-His-Met procarboxypeptidase B coding sequence.
 XX
 KW Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
 KW human serum albumin; premating factor alpha; mating factor alpha;
 KW proCB; ds.
 XX
 OS Sus scrofa.
 XX
 FH Key Location/Qualifiers
 FT CDS
 FT 1..1215
 FT /*tag= a
 FT /product= Porcine procarboxypeptidase B.
 XX
 PN WO9514096-A1.
 XX
 XX 26-MAY-1995.
 XX 16-NOV-1994; 94WO-US13142.
 XX 16-NOV-1993; 93US-0153258.
 XX (ELIL) LILLY & CO ELI.
 XX Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
 PI Sterner JL, Zhang H;
 XX WPI; 1995-200386/26.
 DR P-PSDB; AAR75131.
 XX
 XX DNA encoding porcine carboxypeptidase B - used for transforming
 PT host cells, partic. Pichia species, for prodn. of the enzyme
 XX

Claim 2; Page 20-21; 34pp; English.

The porcine carboxypeptidase B coding sequence can be placed in a bacterial or pref. *Pichia* yeast expression vector. The expression vector further comprises the signal peptide of either human serum albumin (designated pLGD23 - NRRL B-21029); premating factor alpha (designated pFJ489 - NRRL B-21028); mating factor alpha (designated pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide, (designated pLGD27 - NRRL B-21027). The method can be used for producing large amounts of porcine carboxypeptidase B and when produced in *Pichia* yeast, the protein does not need solubilisation or folding. The produced enzyme is then used for pref. cleaving basic residues from the carboxy terminus of proteins.

Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;

Alignment Scores:

Pred. No.:	9, 21e-57	Length:	1215
Score:	582.00	Matches:	117
Percent Similarity:	59.47%	Conservative:	40
Best Local Similarity:	44.32%	Mismatches:	67
Query Match:	43.50%	Indels:	40
DB:	16	Gaps:	5

US-09-980-881A-4 (1-246) x AAQ90600 (1-1215)

Qy	6	TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg	25
Db	310	TATGAGAAGTACAAACAATCGGAAACCATCGAGGCTTGGACTAAGCAAGTACCACAGTGAA	369
Qy	26	HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr	45
Db	370	AATCCAGACCTCATCTCTCGACAGCATTGGAATCATATTTTAGGAAACAATATATATAC	429
Qy	46	ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly	65
Db	430	CTCCTCAAGGTT--GGCAAACTGGACCAAAATAAGCTGCCATTTTCATGGACTGTGGT	486
Qy	66	IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle	81
Db	487	TTCCATGCCAGAGAATGGATTTCCCATGTCATTTTGCAGTGGTTTGTGAGAGAGGCTGTT	546
Qy	81	-----GlyHis-----	81
Db	547	CTCACCTATGGATATGAGAGTCAATGACAGAAATTCCTCAACAAGCTAGACTTTTATGTC	606
Qy	82	-----GlyHis-----AsnArgMetTrpArg	88
Db	607	TTGCCTGTGCTCAATATATGATGGCTCATCTACACCTGGACCAAGAACCGAATGTGAGA	666
Qy	89	LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe	108
Db	667	ARGACCGCTCTACCAATGTGGAACTACCTGCAATTGGCAGACCCCAACAGAAATTTT	726
Qy	109	AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys	128
Db	727	--GATGCTGGTGGTGCACAACTGGAGCCTCTACAGACCCCTCGCATGGACTTACTGT	783
Qy	129	GlyLeuTyrProGluSerGluProGluValLysAlaValaSerPheLeuArgArgAsn	148
Db	784	GGATCTGCTGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGTAATTTTATACGCAACAAC	843
Qy	149	IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro	168
Db	844	CTCTCTCCATCAAAGCATACCTGACCATCCACTCATCTCACAGATGATACTCTACCCCT	903
Qy	169	TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValaSerGlu	188
Db	904	TATTCCTATGATTACAACCTCCCGAGAGCAAACTGCTGATGTGATTAACCTGGCTAAGGCT	963
Qy	189	AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer	208
Db	964	GCCTGTAAGAAGACTT--GCTACACTGTATGGCAACCAAGTACATACGCGCCCAAGGACT	1020

Query Match:	42.38%	Indels:	40
DB:	17	Gaps:	5
US-09-380-881A-4 (1-246) x AAT42497 (1-1053)			
Qy	1	AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAenGluIleTyrSerTrrlleGlu	20
Db	67	GCAACCTGGTCACTCTTACGAGAGTACAAAGTGGGAAACGATAGAGGCTTGGACTCAA	126
Qy	21	PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu	40
Db	127	CAAGTCGCCACTGAGAAATCCAGCCCTCATCTCCGAGTGTATTCGGAACCAATTGAG	186
Qy	41	LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAenAlaIle	60
Db	187	GGACGGCGCTATTACCTCTCGAAGTT---GGCAAAGCTGGACAAAAATAAGCGCTGCCATT	243
Qy	61	TrrlleAspCysGlyIleHisAlaArgGluTrrlleSerProAlaPheCysLeuTrrPhe	80
Db	244	TTCAATGGACTGTGGTTTCATCCAGAGAGTGGATTTCTCTCGCATTTCTGCCAGTGGTTT	303
Qy	81	Ile-----	81
Db	304	GTAAGAGAGCGCTTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAG	363
Qy	82	-----GlyHis-----	83
Db	364	TTAGACTTTTATGTCCTGCTGCTCAATATTGATGGCTACATCTACCTGACCAAG	423
Qy	84	AsnArgMetTrrArgLysAsnArgSerPheTyrAlaAenAsnHisCysIleGlyThrAsp	103
Db	424	AGCCGATTTTGGAGAAAGACTCGCTCCACCATACTGGATCTAGCTCATTTGGCACAGAC	483
Qy	104	LeuAsnArgAsnPheAlaSerLysHisTrrPcysGluGluGlyAlaSerSerSerCys	123
Db	484	CCCAACAGAAATTTT---GATGCTGGTTGGTGGAAATGGAGCTCTCGAAACCCCTGT	540
Qy	124	SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer	143
Db	541	GATGAACCTTACTGTGGACCTCGCCGACAGTCTGAAAGGAGACCAAGGCCCTGGCTGAT	600
Qy	144	PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln	163
Db	601	TTCAATCCGACAAACAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCAA	660
Qy	164	HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer	183
Db	661	ATGATGATCTACCTTACTCATATGCTTACAACTCGGTGAGAACAAATGCTGAGTTGAAT	720
Qy	184	LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr	203
Db	721	GCCCTGGCTAAAGCTACTGTGAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA	777
Qy	204	HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrrlleTyr	223
Db	778	TATGGCCGGGAGACTACAACTATCTCTGCTGGGGGCTCTGACGACTGGGCTTAT	837
Qy	224	AspLeuGlyIleLysTyrSerPheThr	232
Db	838	GACCAAGGAATCAGATATTCTCTTACC	864
RESULT 14			
AAT42494			
ID	AAT42494 standard; DNA; 1263 BP.		
XX	AAT42494;		
AC			
XX			
DT	12-FEB-1997 (first entry)		
XX			
DE	Human pancreatic carboxypeptidase B encoding sequence.		
XX			
KW	ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADBPT;		
KW	mustard-ribonucleotide; antibody directed enzyme prodrug therapy;		

XX	anti-neoplastic; prodrug; reverse polarity; ion pair interaction; reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase; Fd; F(ab') ₂ ; FelB; leader; human carboxypeptidase B; ss.	
XX	Homo sapiens.	
XX		
XX	Key	Location/Qualifiers
XX	CDS	1..1248
FT		/*tag= a
FT		/product= preproenzyme
FT		1..39
FT	sig_peptide	/*tag= b
FT		/note= "partial enzyme secretion leader sequence (pre-sequence); residues -108 to -96"
FT	sig_peptide	40..324
FT		/*tag= c
FT		/note= "pro-sequence; residues -95 to -1"
FT	mat_peptide	385..1245
FT		/*tag= d
FT		/note= "mature enzyme; encodes residues +1 to +307"
XX	W09620011-A1.	
XX	PN	
XX	PD	04-JUL-1996.
XX		
XX	21-DEC-1995;	95WO-CB02991.
XX		
XX	16-AUG-1995;	95GB-0016810.
PR	23-DEC-1994;	94GB-0026192.
XX		
XX	(ZENE) ZENECA LTD.	
XX		
PI	Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;	
PI	Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;	
PI	Tarragona-Fiol A, Taylorson CJ;	
XX		
DR	WPI; 1996-321650/32.	
DR	P-PSDB; AAW06172.	
XX		
FT	Two component system for anti-tumour therapy - comprising targeting moiety linked to mutated enzyme which can transform an anti-neoplastic prodrug	
XX		
PS	Reference Example 14; Page 131; 182pp; English.	
XX		
CC	A two-component system for anti-tumour therapy comprises a targeting moiety linked to a mutated enzyme which can transform an anti-neoplastic prodrug. The system is based on antibody directed enzyme prodrug therapy (ADEPT) using non-naturally occurring mutant forms of host enzymes, pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).	
CC	Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can be used. The present sequence encodes the native prepro-HCPB (see AAW06172) contained in vector pIC11698.	
XX		
SQ	Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;	

Alignment Scores:		
Pred. No.:	5.18e-55	Length:
Score:	567.00	Matches:
Percent Similarity:	58.36%	Conservative:
Best Local Similarity:	41.26%	Mismatches:
Query Match:	42.38%	Indels:
DB:	17	Gaps:
		5
		1363

US-09-980-881A-4 (1-246) x AAT42494 (1-1263)

Qy	1	AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu	20
Db	325	GCAACGAGGACACAGTTATGAGAAGTACAAACAGTGGGAACCATAGAGGGTTGGACTCAA	384
Qy	21	PheIleThrGluArgHisProAspMetLeuThrIleHisIleGlySerSerPheGlu	40
Db	385	CAAGTCGCCACTGAGAATCCAGCCCTCATCTTCGCACTGTTATCGGAACACACATTTTGG	444

DE Human pancreatic carboxypeptidase B encoding sequence, **xx**

XX ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;

QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
 Db 445 GGACGGCTATTACCTCTGAGGTT--GGCAAGCTGGACAAATAAGCTTGCATT 501
 QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
 Db 502 TTCATGGACTGTGTTTCCATGCCAGAGAGTGGATTCTCTGCAATTCGCCAGTGGTTT 561
 QY 81 Ile----- 81
 Db 562 GTAAGAGAGCTGTTCTGCTACCTATGAGCTGAGATCCAACTGACAGAGCTTCTCGACAAG 621
 QY 82 -----GlyHis----- 83
 Db 622 TTAGACTTTTATGCTGCTGCTGCTCAATATTGATGGCTACATCTACCTGACCAAG 681
 QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
 Db 682 AGCCGATTTTGGAGAAAGACTCGCTCCACCACATCTGATCTAGCTGCTGCTGACAGAC 741
 QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
 Db 742 CCCAACAGAAATTT---GATGCTGGTGTGTGTAATTTGGAGCTCTCGAAACCCCTGT 798
 QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
 Db 799 GATGAACCTTACTGTGAGCTCGCGCAGACTCTGAAAGAGACCAAGCCCTGGCTGAT 858
 QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
 Db 859 TTCATCGCAACAACCTCTCTCCATCAAGGATATCTGACATCCACTGCTACTCCCA 918
 QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
 Db 919 ATGATGATCTACCTTACTCATCTATGCTTACAAACTCGGTGAGAACAAATGCTGAGTTGAAT 978
 QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
 Db 979 GCCTGGCTAAGCTACTGTGAAGAAGCTT---GCCTCACTGACCGCACCAAGTACACA 1035
 QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
 Db 1036 TATGGCCGGGAGCTACAACAATATCTGCTGCTGGGGGCTCTGACGAGTGGCTTAT 1095
 QY 224 AspLeuGlyIleLysTyrSerPheThr 232
 Db 1096 GACCAAGGAATCAGATATCTCTTCAACC 1122
 RESULT 15
 AAV41795
 ID AAV41795 standard; DNA; 1263 BP.
 AC AAV41795;
 XX
 DT 20-NOV-1998 (first entry)
 XX
 DE Human pancreatic carboxypeptidase nucleotide sequence.
 XX
 KW ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
 KW prodrug therapy.
 XX
 OS Homo sapiens.
 XX
 Key Location/Qualifiers
 FT CDS 1..1248
 FT /*tag= a
 FT /product= "Pancreatic carboxypeptidase B"
 FT /note= "No start codon given"
 XX
 PN W09835988-A1.
 XX
 PD 20-AUG-1998.

XX 10-FEB-1998; 98WO-GB00415.
 PF
 XX 29-OCT-1997; 97GB-0022727.
 PR 14-FEB-1997; 97GB-0003104.
 PR 18-OCT-1997; 97GB-0022003.
 XX
 PA (ZENE) ZENEPA (ZENE) ZENECA LTD.
 XX
 XX Edge MD;
 PI
 XX WPI; 1998-467168/40.
 DR P-PSDB; AAW74476.
 DR
 XX
 PT New modified pro-domain of carboxy-peptidase B - enhances expression
 of co-expressed proteins for production of recombinant
 PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
 PT enzyme prodrug therapy
 XX
 PS Example 1; Page 52-53; 83pp; English.
 XX
 CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
 modified pro-domain of CPB on a separate gene to enhance recombinant
 CC expression. This process can be used to produce recombinant CPB in
 CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
 CC insulin production and protein sequencing, while its fusions with
 CC antibodies are useful in antibody-directed enzyme prodrug therapy. The
 CC modified pro-domain provide increased yields of recombinant CPB, possibly
 CC by protecting the C-terminus against enzymatic degradation or by
 CC increasing intracellular trafficking.
 XX
 SQ Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;
 Alignment Scores:
 Pred. No.: 5,18e-55 Length: 1263
 Score: 567.00 Matches: 111
 Percent Similarity: 58.36% Conservative: 46
 Best Local Similarity: 41.26% Mismatches: 72
 Query Match: 42.38% Indels: 40
 DB: 19 Gaps: 5
 US-09-980-881A-4 (1-246) x AAV41795 (1-1263)
 QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
 Db 325 GCACAGGACACAGTGTATGAGAAGTACAAACAGTGGGAAACGATAGAGCTTGGACTCAA 384
 QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
 Db 385 CAACTGCCCACTGAGATCCAGCCCTCATCTCTCGCAGTGTATCGGAACCACTTTGAG 444
 QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
 Db 445 GGACGGCTATTACCTCTGAGGTT--GGCAAGCTGGACAAATAAGCTTGCATT 501
 QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
 Db 502 TTCATGGACTGTGTTTCCATGCCAGAGAGTGGATTCTCTGCAATTCGCCAGTGGTTT 561
 QY 81 Ile----- 81
 Db 562 GTAAGAGAGCTGTTCTGCTACCTATGAGCTGAGATCCAACTGACAGAGCTTCTCGACAAG 621
 QY 82 -----GlyHis----- 83
 Db 622 TTAGACTTTTATGCTGCTGCTGCTCAATATTGATGGCTACATCTACCTGACCAAG 681
 QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
 Db 682 AGCCGATTTTGGAGAAAGACTCGCTCCACCACATCTGATCTAGCTGCTGCTGACAGAC 741
 QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
 Db 742 CCCAACAGAAATTT---GATGCTGGTGTGTGTAATTTGGAGCTCTCGAAACCCCTGT 798
 QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
 Db 799 GATGAACCTTACTGTGAGCTCGCGCAGACTCTGAAAGAGACCAAGCCCTGGCTGAT 858
 QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
 Db 859 TTCATCGCAACAACCTCTCTCCATCAAGGATATCTGACATCCACTGCTACTCCCA 918
 QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
 Db 919 ATGATGATCTACCTTACTCATCTATGCTTACAAACTCGGTGAGAACAAATGCTGAGTTGAAT 978
 QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
 Db 979 GCCTGGCTAAGCTACTGTGAAGAAGCTT---GCCTCACTGACCGCACCAAGTACACA 1035
 QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
 Db 1036 TATGGCCGGGAGCTACAACAATATCTGCTGCTGGGGGCTCTGACGAGTGGCTTAT 1095
 QY 224 AspLeuGlyIleLysTyrSerPheThr 232
 Db 1096 GACCAAGGAATCAGATATCTCTTCAACC 1122
 RESULT 15
 AAV41795
 ID AAV41795 standard; DNA; 1263 BP.
 AC AAV41795;
 XX
 DT 20-NOV-1998 (first entry)
 XX
 DE Human pancreatic carboxypeptidase nucleotide sequence.
 XX
 KW ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
 KW prodrug therapy.
 XX
 OS Homo sapiens.
 XX
 Key Location/Qualifiers
 FT CDS 1..1248
 FT /*tag= a
 FT /product= "Pancreatic carboxypeptidase B"
 FT /note= "No start codon given"
 XX
 PN W09835988-A1.
 XX
 PD 20-AUG-1998.

```
Db 742 CCCAACAGAAATTT---GATGCTGGTGTGTGAATTTGGAGCCTCTCGAAACCCCTGT 798
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 799 GATGAAACTTACTGTGGACCTGCCGAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGAT 858
QY 144 PheLeuArgArgAsnIleAenGlnIleLeuAlaTyrIleSerMethHisSerTyrSerGln 163
Db 859 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 918
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 919 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAGAACAAATGCTGAGTTGAAT 978
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 979 GCCCTGGCTAAAGCTACTGTGAAGAAGACTT--GCCTCACTGCACGGCACCAAGTACACA 1035
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrolleTyr 223
Db 1036 TATGGCCCGGAGCTACAAACAATCTATCTGTGCTGGGGGCTCTGACGACTGGGCTTAT 1095
QY 224 AspLeuGlyIleLysTyrSerPheThr 232
Db 1096 GACCAAGGAATCAGATATTCCTTCACC 1122
```

Search completed: January 1, 2004, 20:09:34
Job time : 279.597 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:51:12 ; Search time 112.985 Seconds
(without alignments)
961.011 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQHSLNEIYSWIE.....IKVFTSPNPVKKLPLSLK 246

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-Qs/cgn2_1/USPTO pool/US0980881/runat_31122003_135917_26894/app_query.fasta_1.1628
-DB=Issued_Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0980881 @CGN 1 1 115 @runat_31122003_135917_26894 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1289.5	96.4	1625	4	US-09-813-133A-1
2	1272	95.1	1272	2	US-08-869-057-1
3	1272	95.1	1749	1	US-07-649-591B-2
4	1272	95.1	1749	1	US-08-277-540-2
5	1272	95.1	1749	1	US-08-430-787A-2
6	610	45.6	927	2	US-08-782-760-5
7	610	45.6	927	5	PCT-US96-00995-5
8	582	43.5	921	1	US-08-696-139-3
9	582	43.5	1215	1	US-08-696-139-1
10	567	42.4	999	2	US-08-860-882A-67
11	567	42.4	999	4	US-09-011-769A-50
12	567	42.4	1053	2	US-08-860-882A-64

13	567	42.4	1053	4	US-09-011-769A-46	Sequence 46, Appl
14	567	42.4	1263	2	US-08-860-882A-56	Sequence 56, Appl
15	567	42.4	1263	4	US-09-011-769A-38	Sequence 38, Appl
16	567	42.4	1284	2	US-08-860-882A-71	Sequence 71, Appl
17	567	42.4	1284	4	US-09-011-769A-55	Sequence 55, Appl
18	560	41.9	1053	4	US-09-463-451-27	Sequence 27, Appl
19	560	41.9	1053	4	US-09-463-451-28	Sequence 28, Appl
20	560	41.9	1059	2	US-08-860-882A-74	Sequence 74, Appl
21	560	41.9	1059	4	US-09-011-769A-59	Sequence 59, Appl
22	559	41.8	1059	2	US-08-860-882A-77	Sequence 77, Appl
23	559	41.8	1059	4	US-09-011-769A-63	Sequence 63, Appl
24	556.5	41.6	1311	4	US-09-675-305-9	Sequence 9, Appl
25	549	41.0	1870	3	US-09-171-945-112	Sequence 112, App
26	549	41.0	2154	3	US-09-171-945-124	Sequence 124, App
27	525.5	39.3	2128	4	US-09-675-305-13	Sequence 13, Appl
28	492	36.8	1200	4	US-09-710-099-7	Sequence 7, Appl
29	471.5	35.2	1050	4	US-09-675-305-11	Sequence 11, Appl
30	465.5	34.8	1257	3	US-08-640-906-1	Sequence 1, Appl
31	465.5	34.8	1257	4	US-09-395-936-1	Sequence 1, Appl
32	465.5	34.8	1311	4	US-09-710-099-5	Sequence 5, Appl
33	459	34.3	1251	3	US-08-640-906-3	Sequence 3, Appl
34	459	34.3	1251	4	US-09-395-936-3	Sequence 3, Appl
35	405.5	30.3	945	4	US-09-710-099-3	Sequence 3, Appl
36	405.5	30.3	945	4	US-09-710-099-11	Sequence 11, Appl
37	403.5	30.2	2247	4	US-09-710-099-15	Sequence 15, Appl
38	379	28.3	1056	4	US-09-710-099-1	Sequence 1, Appl
39	379	28.3	1056	4	US-09-710-099-9	Sequence 9, Appl
40	356	26.6	55827	4	US-09-813-133A-3	Sequence 3, Appl
41	225.5	16.9	741	4	US-09-675-305-5	Sequence 5, Appl
42	198.5	14.8	629	4	US-09-280-116-228	Sequence 228, App
43	169.5	12.7	494	4	US-09-669-751-63	Sequence 63, Appl
44	169.5	12.7	515	3	US-08-998-416-125	Sequence 125, App
45	163	12.2	591	3	US-09-331-709-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-1
; Sequence 1, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-1

Alignment Scores:
Pred. No.: 5.61e-153 Length: 1625
Score: 1289.50 Matches: 244
Percent Similarity: 92.42% Conservative: 0
Best Local Similarity: 92.42% Mismatches: 2
Query Match: 96.38% Indels: 18
DB: 4 Gaps: 1

US-09-980-881A-4 (1-246) x US-09-813-133A-1 (1-1625)

QY	1	AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPileGlu	20
DB	359	GCCTCCGATCGTACTATGACAGTATCACTCACTAAGTAATCTATTTCTTGATAGAA	418
QY	21	PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu	40

Db 419 TTTATAACTGAGGCGATCCTGATATGCTTACAAAAATCCACATTCGATCCTCATTTGAG 478
Qy 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluThrAlaLysAsnAlaIle 60
Db 479 AAGTACCACCTCTATGTTTAAAGGTTTCTGGAAGAAACAAGCAGCCAAAAATGCCATA 538
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 539 TGGATTGACTGTGAATCCATGCCAGAGATGATCTCTCTGCTTCTCTGTTGTGGTTC 598
Qy 81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIle 100
Db 599 ATAGGCCATAATCAATGTGGAGAAAGAACCGTTCTTCTATGCGAACATCATTTGCATC 658
Qy 101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
Db 659 GGAACAGACCTGAATAGGAATTTGCTTCCAAACACATGGTGTGAGGAAGTGCATCCAGT 718
Qy 121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
Db 719 TCCTCATGCTCGGAAACCTACTGTGGAATTTATCCTGAGTCAGAACAGAGTGAAGGCA 778
Qy 141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
Db 779 GTGGCTAGTTTCTTGAGAGAAATATCAACAGATTAAAGCATACATCAGCATGCATTCA 838
Qy 161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
Db 839 TACTCCAGCATATAGTGTTCATATATTCCTATACAGCAAGTAAAGCAAGACCATGAG 898
Qy 181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
Db 899 GAACGTCTCTAGTAGCAGTAGAGCAGTTCGCTGCTATTGAGAAATATTAGTAAATAACC 958
Qy 201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220
Db 959 AGGTATACATGATGCCATGCTGCTCAGAAACCTTATACCTAGCTCTCTGGAGTGGGACGAT 1018
Qy 221 TrpIleTyrAspLeuGlyIleLysTyrSerPhe----- 231
Db 1019 TGGATCTATGATTGGGCACTCAATATTCGTTTACAATTGAACCTCGAGATACGGGCACA 1078
Qy 232 -----ThrSerAsnProProValGluLysLeuLeuProLe 243
Db 1079 TAGGATCTTGTGCGGAGCGGTATACATCAACCCACCTGTAGAGAAGCTTTTGGCGCT 1138
Qy 243 uSerLeuLys 246
Db 1139 GTCTCTAAAA 1148

RESULT 2

US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Washtien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-5411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PUBLICATION INFORMATION:
; AUTHORS: Eaton, Dan L.
; AUTHORS: Malloy, Beth E.
; AUTHORS: Tsai, Siao P.
; AUTHORS: Henzel, William
; AUTHORS: Drayna, Dennis
; TITLE: Isolation, Molecular Cloning, and Partial
; TITLE: Characterization of a No. 5985562el Carboxypeptidase B
; TITLE: from Human Plasma
; JOURNAL: J. Biol. Chem.
; VOLUME: 266
; ISSUES: 32
; PAGES: 21833-21838
; DATE: No. 5985562 15-1991
; US-08-869-057-1

Alignment Scores:

Pred. No.: 6,17e-151 Length: 1272
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-4 (1-246) x US-08-869-057-1 (1-1272)

Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 343 GCCTCGCATCGTACATGATGACAGATGATCTACTAAATGAAATCTATTCTTGATAGAA 402
Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 403 TTTATACTGAGAGGATCCTGATATGCTTACAAAAATCCACATTCGATCCTCATTTGAG 462
Qy 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 463 AAGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAACAACAGCAAAATGCCATA 522
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 523 TGGATTGACTGTGGAATGCCATGCCAGAAATGGAATCTCTCTCTCTCTCTCTCTCTCT 582
Qy 81 IleGlyHis----- 83
Db 583 ATAGGCCATATACTCAATCTATGGGTAATAGGCAATATACCAATCTCTCTGAGCCTT 642
Qy 83 ----- 83
Db 643 GTGGATTTCTATGTTATGCGGTTGTTAATGTGCGGTTATGCGGTTATGACTACTCATGAAAAAG 702
Qy 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 103
Db 703 AATCGAATGTGGAAGAACCGTTCTTCTATGCGAACAAATCATTCGATCGGAACAGAC 762
Qy 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123

Db 763 CTGAATAGGAACCTTCTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGC 822
Qy 124 SerGluThrTyCysGlyLeuTyProGluSerGluProGluVallyAlaValAlaSer 143
Db 823 TCGAAACCTACTGTGGACTTTATCTGAGTCAGAACCAAGTGAAGCGAGTGGCTAGT 882
Qy 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyIleSerMethHisSerTySerGln 163
Db 883 TTCTTGAGAGAAATATCAACCAAGATTAAGCATATCATGACATGATTCATCTCCAG 942
Qy 164 HisIleValPheProTySerTyThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 943 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGCAACCATGAGGAACGTCT 1002
Qy 184 LeuValAlaSerGluAlaValArgAlaIleGluTyThrSerLysAsnThrArgTyThr 203
Db 1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAACACTAGTAAAAATACCAAGTATACA 1062
Qy 204 HisGlyHisGlySerGluThrLeuTyLeuAlaProGlyGlyValAspAspTyrPhe 223
Db 1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGACGATTGGATCTAT 1122
Qy 224 AspLeuGlyIleLysTySerPhe----- 231
Db 1123 GATTGGGCATCAAAATTCGTTTACAATTGAATTCGAGATACGGGCACATACGATTC 1182
Qy 232 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeu 246
Db 1183 TTGCTCCGCGAGGCTTACATCAACCCACTGTGAGAGAGCTTTTCCGCTGTCTCTAAA 1242
Qy 246 s 246
Db 1243 A 1243

RESULT 3

US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID

STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-07-649-591B-2
Alignment Scores:
Pred. No.: 1.03e-150 Length: 1749
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 1 Gaps: 2
US-09-980-881A-4 (1-246) x US-07-649-591B-2 (1-1749)
Qy 1 AlaSerAlaSerTyTyGluGlnTyHisSerLeuAsnGluIleTySerTyrPheGlu 20
Db 383 GCCTCGCATCTGCTATCATGAAACAGATCATCTCACTAAATGAATCTATTCTTGGATAGAA 442
Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 443 TTTATAACTGAGAGGCATCTGATATGCTTACAAAATCCAAATCGATTCCTCATTTGAG 502
Qy 41 LysTyProLeuTyValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 503 AAGTACCACCTCTATGTTTTAAAGGTTTCTGGAAAGAAACAAACGCAAAATATGCATA 562
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTyrPhe 80
Db 563 TGGATTGACTGTGGAATCCATGCCAGAAATGATCTCTCTGCTTTCTGCTTTGCTGTTTC 622
Qy 81 IleGlyHis----- 83
Db 623 ATAGGCCATATACTCAATCTATGGGATAATAGGCAATATACCAATCTCTCTGAGGCTT 682
Qy 83 ----- 83
Db 683 GTGGATTTCTATGTTATGCGGTGGTTAATGTGGACGGTTATGACTACTCATGAAAAAG 742
Qy 84 AsnArgMetTyrArgLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAsp 103
Db 743 AATCGAATGTGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTCGATCGAACAGAC 802
Qy 104 LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGlyValaSerSerSerCys 123
Db 803 CTGAATAGGAACCTTTCCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862
Qy 124 SerGluThrTyCysGlyLeuTyProGluSerGluProGluVallyAlaValAlaSer 143
Db 863 TCGAAACCTACTGTGGACTTTATCTGAGTCAGAACCAAGTGAAGCGAGTGGCTAGT 922
Qy 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyIleSerMethHisSerTySerGln 163
Db 923 TTCTTGAGAGAAATATCAACCAAGATTAAGCATATCATGACATGATTCATCTACTCCCAG 982
Qy 164 HisIleValPheProTySerTyThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 983 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGCAACCATGAGGAACGTCTCT 1042

184 LeuValalaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
1043 CTAGTAGCAGTGAAGCAGTTCGTCTATTGAGAACTAGTAAATACAGGTATACA 1102
204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspThrPileTyr 223
1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGACGATTGGATCTAT 1162
224 AspLeuGlyIleLysTyrSerPhe----- 231
1163 GATTGGGCATCAAAATTCGTTTACAAATGAACTTCGAGATACGGGCACATACGGATTC 1222
232 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLys 246
1223 TTGCTGCCGAGCGTTACATCAAAACCCTGTAGAGAGCTTTTGGCGGTGCTCTTAA 1282
246 s 246
1283 A 1283

RESULT 4
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Haasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1CID1
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-277-540-2

Alignment Scores:
Pred. No.: 1.03e-150 Length: 1749
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0

Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 1 Gaps: 2

US-09-980-881A-4 (1-246) x US-08-277-540-2 (1-1749)

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPileGlu 20
Db 383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAATGAATCAATCTTCTTGGATAGAA 442
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 443 TTTATAACTCAGAGGCATCTGATATGCTTACAAAAATCCACATTCGATCCTCATTTGAG 502
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 503 AAGTACCCACTCTATGTTTAAAGGTTTCTGGAAAAAGAACAAACAGCCAAAAATGCCATA 562
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPileSerProAlaPheCysLeuTyrPhe 80
Db 563 TGGATTGACTGTGGAAATCCATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGGTTC 622
QY 81 IleGlyHis----- 83
Db 623 ATAGGCCATATAACTCAATTCATGGGATAATAGGCAATATATACCAATCTCTCTGAGCCT 682
QY 83 ----- 83
Db 683 GTGGATTTCTATGTTATGCGGTGTTAATGTGCGGGTATGACTACTCATGCGAAAAAG 742
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 743 AATCGAATGTGGAGAAAGAACCGTCTTCTATGCGACATCATTCATCGAACAGAC 802
QY 104 LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerCys 123
Db 803 CTGAATAGGAACCTTTCCTTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGC 862
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 863 TCGAAAAACCTACTGTGGACTTTATCTGAGTCAAGAACCAAGAGTGAAGCGAGTGGCTAGT 922
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 923 TTCTTGAGAGAAATATCAACAGATTAAGCATATACATCAGCATGCTCATCTACTCCAG 982
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 983 CATATAGTGTTCATATTCCTATATACAGAAAGTAAAAAGCAAGACCATGAGGAACGTCT 1042
QY 184 LeuValalaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1043 CTAGTAGCCAGTGAAGCAGTTCGTCTATTGAGAAAACTAGTAAAAATACAGGTATACA 1102
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspThrPileTyr 223
Db 1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGACGATTGGATCTAT 1162
QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 1163 GATTGGGCATCAAAATTCGTTTACAAATGAACTTCGAGATACGGGCACATACGGATTC 1222
QY 232 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLys 246
Db 1223 TTGCTGCCGAGCGTTACATCAAAACCCTGTAGAGAGCTTTTGGCGGTGCTCTTAA 1282
QY 246 s 246
Db 1283 A 1283

RESULT 5
US-08-430-787A-2
; Sequence 2, Application US/08430787A

; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Baton, Dan L.
; TITLE OF INVENTION: No. 5593674e1 Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 350 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430, 787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277, 540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-430-787A-2

Alignment Scores:
Pred. No.: 1.03e-150 Length: 1749
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 1 Gaps: 2

US-09-980-881A-4 (1-246) x US-08-430-787A-2 (1-1749)

QY 1 AlaSerAlaSerTyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 20
DB 383 GCTTCCGCGATCGTACTATGACAGATGATCACTCAATAAGAAATCTATTCTCGGATAGAA 442
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 443 TTTTAACTGAGAGGCATCTCTGATATGCTTCAAAAATCCACATTTGATCTCTATTGAG 502
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 503 AAGTACCACCTCTATGTTTAAAGGTTTCTGGAAGAAACAAACAGCAAAATGCCATA 562
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
DB 563 TGGATTGACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTTGTGTTTC 622

QY 81 IleGlyHis----- 83
DB 623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGCAATATACCAATCTCTCGAGGCTT 682
QY 83 ----- 83
DB 683 GTGGATTCTATGTTATGCGGGTGGTTAATGTGGAGGTTATGACTACTCATCGAANAAG 742
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 103
DB 743 AATCGAATGTGGAGAAAGAACCGTCTTCTTATGCGAACCAATCATTCGATCGGAACAGAC 802
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluClyAlaSerSerSerCys 123
DB 803 CTGAATAGGAACCTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTCTCTCATGC 862
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
DB 863 TCGAAACCTTACTGTGGACTTTATCTGAGTCAGAACCAAGAGGAGGAGGAGTGGCTAGT 922
QY 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyrIleSerMethHisSerTyrSerGln 163
DB 923 TTCTTCGAGAGAAATATCAACCAGATTAAAGCATATCATCAGCATGCATTCATATCCAG 982
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
DB 983 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGCAACCATGAGGAACCTGTCT 1042
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
DB 1043 CTAGTAGCCAGTGAAGCAGTTCGTCTATTGAGAAAACCTAGTAAATAATACCAAGTATACA 1102
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTyrIleTyr 223
DB 1103 CATGGCCATGGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGACGATGGAATCTAT 1162
QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
DB 1163 GATTTCGGCATCAATATTCGTTTACAAATTGAACCTTCGAGATACGGGCACATACGGATTC 1222
QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
DB 1223 TTGCTCGCGAGGCGTTTACATCAAAACCCACCTGTAGAGAAAGCTTTTCCGCTGTCTCTAAA 1282
QY 246 s 246
DB 1283 A 1283

RESULT 6
US-08-782-760-5
; Sequence 5, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:
; APPLICANT: Hartman, Jacob
; APPLICANT: Fulga, Netta
; APPLICANT: Mendelovitch, Simona
; APPLICANT: Gorecki, Marian
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30


```

CURRENT APPLICATION DATA:
APPLICANT: US/08/782,760
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/378,233
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0336/43847
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 927 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..927
US-08-782-760-5

Alignment Scores:
Pred. No.: 2,05e-67
Score: 610.00
Percent Similarity: 60.50%
Best Local Similarity: 44.84%
Query Match: 45.59%
DB: 2
Gaps: 5

US-09-980-881A-4 (1-246) x US-08-782-760-5 (1-927)

QY 1 AlaSerAlaSerTyrTyrGluGluTyrHisSerLeuAsnGluLeuTyrSerTrpIleGlu 20
DB 1 GCAAGTGCACACAGCTACACCAAGTACCAACATGGGAAACGATTGAGCGTGGATTCAA 60
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 61 CAAGTTCGCACTGATATATCAGACCTTGCTACTCAGAGCGTCATTGGAACCAATTGAA 120
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 121 GGACGTAACTATGTCTCAAGATT--GGTAAAACTAGACCGCAATAAGCGCTGCCATC 177
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
DB 178 TTCATGATTGTGGTTCATGCAAGAGAGTGGATTCTCTGCAATCTCTGAGTGGTTT 237
QY 81 Ile-----81
DB 238 GTGAGAGAGGCTGTCGCTACCTATATCAAGAGATCCACATGAACAGCTTCTAGATGAA 297
QY 82 -----GlyHis-----83
DB 298 CTGGATTCTATGTTCTGCTGTGGTCAACATTGATGGCTATGCTACACTGAGCTAAG 357
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 103
DB 358 GACAGAAATGTGGAAAAACCCGCTCTACTATGCTGGAAGTTCTCTGCTGGGTGTAGAC 417
QY 104 LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerCys 123
DB 418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGAGCTTCTCGAGTCCCTGC 474
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
DB 475 TCTGAAACTTACTGTGGACGAGCCCGAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

```

144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
535 TTCTATCGCAACCACTCTCCACCATCAGGCTACTCTGACCATCCTACTACTACAG 594

164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
595 ATGATGCTCTACCTTACTCTATGACTACAAACTGCTGAGAACTATGAGGAATTGAAT 654

184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
655 GCCTGTGTGAAAGGTGCGCAAGAGCTT--GCCACTCTGTCATGCGCAACCAAGTACACA 711

204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
712 TATGGCCAGGAGCTACCAACATCTATCTGCTGCTGGGGATCTGACGACTGCTTAT 771

224 AspLeuGlyIleLysTyrSerPheThrSerAsnProValGluLysLeuLeuProLeu 243
772 GATCAGGGAATCAATATTCTTTACCTTTTGAACCT--CCGGGATACAGGCTTCTTTGGCTT 830

244 Ser 244
831 TCT 833

RESULT 7
PCT-US96-00995-5
Sequence 5, Application PC/TUS9600995
GENERAL INFORMATION:
APPLICANT: Bio-Technology General Corp.
TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
TITLE OF INVENTION: CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00995
FILING DATE: 25-JAN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0336/43847-A-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 927 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..927
PCT-US96-00995-5

Alignment Scores:
Pred. No.: 2,05e-67
Score: 610.00
Matches: 126
Conservative: 44
Mismatch: 71
Indels: 41
Gaps: 5

Percent Similarity: 60.50% Conservatives: 44
Best Local Similarity: 44.84% Mismatches: 71
Query Match: 45.59% Indels: 41
DB: 5 Gaps: 5

US-09-980-881A-4 (1-246) x PCT-US96-00995-5 (1-927)

```
QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
DB 1 GCAAGTGGACAGCTACCAAGTACCAACAGTGGAAACAGTGGCGTGGATTCAA 60
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 61 CAGTTGCCACTATATACGACCTGTCTCAGAGCGTCAATGGAACCACTTTGAA 120
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 121 GGACGTAAACATGATGTCTCTCAAGATT--GGTAAAACTAGACCGAATGAAGCTGCCATC 177
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
DB 178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTCTCTCGCATTTCTGTCAAGTGGTTT 237
QY 81 Ile----- 81
DB 238 GTGAGAGAGCTGCTCGTACCTATATCAAGAGATCCATGAACAGCTTCTAGATGAA 297
QY 82 -----GlyHis----- 83
DB 298 CTGGATTCTATGTTCTGCTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
DB 358 GACAGATGTGGAGAAAACCCGCTCTACTATGCTGGAGTTCCTGCTGGGTGTAGAC 417
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluCluAlaSerSerSerCys 123
DB 418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGAGCTTCTCGAGTCCCTGC 474
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
DB 475 TCTGAAACTTACTGTGGACAGCCCGACAGTCTGAAAGAGACAAAGCCCTGGCAGAT 534
QY 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
DB 535 TTCATCGCAACCAACCTCTCCACCATCAGGCTTACTGACCATCCACTCCTACTCAGAG 594
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
DB 595 ATGATGCTCTACCTTACTCTCTATGACTACAAACTGCTGAGAACTATGAGGAATGAAT 654
QY 184 LeuValAlaSerGluAlaValAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
DB 655 GCCTGTGTGAAGGTGGGCAAGGAGCTT---GCCACTGTGATGGCACCAGATACACA 711
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
DB 712 TATGGCCAGAGCTACAACTATCTCTGCTGCTGGGGGATCTGACGACTGCTCTTAT 771
QY 224 AspLeuGlyIleLysTyrSerPheThrSerAsnProValGluLysLeuLeuProLeu 243
DB 772 GATCAGGAATCAAAATATCTCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830
QY 244 Ser 244
DB 831 TCT 833
```

RESULT 8

US-08-696-139-3

; Sequence 3, Application US/08696139

; Patent No. 5672496

; GENERAL INFORMATION:

; APPLICANT: Fayerman, Jeffrey T.

; APPLICANT: Greenen, David P.
; APPLICANT: Hersberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Steiner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gavlo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..921
; US-08-696-139-3

Alignment Scores:

Pred. No.: 7,07e-64 Length: 921
Score: 582.00 Matches: 117
Percent Similarity: 59.47% Conservative: 40
Best Local Similarity: 44.32% Mismatches: 67
Query Match: 43.50% Indels: 40
DB: 1 Gaps: 5

US-09-980-881A-4 (1-246) x US-08-696-139-3 (1-921)

```
QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
DB 16 TATGAGAAGTACCAACACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTACCAGTGAA 75
QY 26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
DB 76 AATCCAGACTCATCTCTCGCACAGCATCGGAACCTACATATTTTAGGAAACAATATATAC 135
QY 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
DB 136 CTCCTCAAGGTT---GGCAAACTGGACCAATAAGCCTGCCATTTTCATGGACTGGT 192
QY 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 81
DB 193 TTCATGCCAGAGATGGATTTCCTCATGCAATTTTTCAGTGGTTGTGAGAGAGGCTGTT 252
QY 81 ----- 81
```

253 CTCACCTATGATATGAGAGTTCACATGACAGAAATTCCTCAACAGCTAGACTTTATGTC 312
QY 82 -----GlyHis-----AsnArgMetTyrArg 88
Db 313 TTGCCTGTCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA 372
QY 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
Db 373 AGAGCCGCTCTACCAATGCTGGAACTACCTGCTGATGGCAGACCCCAACAGAAATTT 432
QY 109 AlaserLysHisTyrCysGluGluGlyAlaserSerSerCysSerGluThrTyrCys 128
Db 433 ---GATCCTGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGGGATGAGACTTACTGT 489
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaserPheLeuArgAsn 148
Db 490 GGATCTGCTGCAGAGTCTGAAAGAGACCAAGCCCTGGCTGATTTTATACGCAACAA 549
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
Db 550 CTCTCTCCATCAAGCATACCTGACGATCCACTCATCTACAGATGATACTTACCCT 609
QY 169 TyrSerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaserGlu 188
Db 610 TATTCCTATGATTACAACTCCCGAGAACAACTGCTGAGTTGAATACTGGCTAAGGCT 669
QY 189 AlavalargAlaleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
Db 670 GCGGTGAAGAACTT---GCTACACTGTATGGCACCAGATACACATACGCGCCAGGAGCT 726
QY 209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIleLys 228
Db 727 ACACAACTATCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAA 786
QY 229 TyrSerPheThr 232
Db 787 TATTCCTTACC 798

RESULT 9

US-08-696-139-1
Sequence 1, Application US/08696139
Patent No. 5672496
GENERAL INFORMATION:
APPLICANT: Fayerman, Jeffrey T.
APPLICANT: Greenen, David P.
APPLICANT: Hershberger, Charles L.
APPLICANT: Larson, Jeffrey L.
APPLICANT: Sterner, Jane L.
APPLICANT: Zhang, Haichao
TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,139
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,258
FILING DATE: 16-NOV-1993
ATTORNEY/AGENT INFORMATION:

NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-8681
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1215 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1215
US-08-696-139-1

Alignment Scores:

Pred. No.:	1,116-63	Length:	1215
Score:	582.00	Matches:	117
Percent Similarity:	59.47%	Conservative:	40
Best Local Similarity:	44.32%	Mismatches:	67
Query Match:	43.50%	Indels:	40
DB:	1	Gaps:	5

US-09-980-881A-4 (1-246) x US-08-696-139-1 (1-1215)

QY	6	TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArg	25
Db	310	TATGAGAGTACAACTGGGAAACGATCGAGCTTGGACTTAAGCAAGTACCAGTGAA	369
QY	26	HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr	45
Db	370	AATCCAGACCTCTCTCTCGACAGCCATCGGAACATCATATTTTAGGAACATATATAC	429
QY	46	ValLeuLysValSerGlyLysGluThrAlaLysAsnAlaIleTyrIleAspCysGly	65
Db	430	CTCCTCAAGTT---GGCAACCTGGACCAATAGCCTGCCATTTTCATGGACTGTGT	486
QY	66	IleHisAlaArgGluTyrIleSerProAlaPheCysLeuTyrPheIle	81
Db	487	TTCCATGCCAGAGATGGATTTCCCATGCATTTTGGCAGTGGTTTGTGAGAGAGGCTGT	546
QY	81	-----GlyHis-----AsnArgMetTyrArg	88
Db	547	CTCACCTATGGATGATGAGATGCATGACAGAAATTCCTCAACAGCTAGACTTTTATGTC	606
QY	82	-----GlyHis-----AsnArgMetTyrArg	88
Db	607	TTGCCTGTCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA	666
QY	89	LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe	108
Db	667	AGAGCCGCTCTACCAATGCTGGAACTACCTGCTGATGGCAGACCCCAACAGAAATTT	726
QY	109	AlaserLysHisTyrCysGluGluGlyAlaserSerSerCysSerGluThrTyrCys	128
Db	727	---GATCCTGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGGGATGAGACTTACTGT	783
QY	129	GlyLeuTyrProGluSerGluProGluValLysAlaValAlaserPheLeuArgAsn	148
Db	784	GGATCTGCTGCAGAGTCTGAAAGAGACCAAGCCCTGGCTGATTTTATACCAACAA	843
QY	149	IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro	168
Db	844	CTCTCTCCATCAAGCATACCTGACGATCCACTCATCTACAGATGATACTTACCCT	903
QY	169	TyrSerTyrThrArgSerLysSerLysAspHisGluLeuSerLeuValAlaserGlu	188
Db	904	TATTCCTATGATTACAACTCCCGAGAACAACTGCTGAGTTGAATTAACCTGGCTAAGGCT	963
QY	189	AlavalargAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer	208


```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-860-882A-64

```

Alignment Scores:	6.93e-62	Length:	1053
Pred. No.:	567.00	Matches:	111
Score:	58.36%	Conservative:	46
Percent Similarity:	41.26%	Mismatches:	72
Best Local Similarity:	42.38%	Indels:	40
Query Match:	2	Gaps:	5
DB:	2		

US-09-980-881A-4 (1-246) x US-08-860-882A-64 (1-1053)

Qy	1	AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLleTyrSerTrpIleGlu	20
Db	67	GCACCTGGTCTACTCTTACGAGAAGTACAAAGTGGGAAACGATAGAGGGCTTGGACTCAA	126
Qy	21	PhelIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu	40
Db	127	CAAGTGGCCACTGAGAAATCCAGCCCTCATCTCTCGCAGTGTATTCGAGACCAATTGTAG	186
Qy	41	LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle	60
Db	187	GGACGCGCTATTACCTCTCTGAAGGTT---GGCAAAAGCTGGACAAAAATAAGCCCTGCCATT	243
Qy	61	TrpIleAspCysGlyIleHisAlaArgLutTrpIleSerProAlaPheCysLeuTrpPhe	80
Db	244	TTCATGGACTGTGGTTTCCATGCCAGAGTGGATTTCTCTGCTATTCTCCAGTGGTTT	303
Qy	81	Ile-----	81
Db	304	GTAAGAGAGCGTGTTCGTACCTATGCGACGTGAGTCAAGTGACAGAGCTTCTCGACAA	363
Qy	82	-----GlyHis-----	83
Db	364	TTAGACTTTTATGCTCGCTGTGCTCAATATTGATGGCTACATCTACACTGGACCAAG	423
Qy	84	AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp	103
Db	424	AGCCGATTTTGGAGAAGAAGACTCGCTCCACCATACTGGATCTAGCTGCATTTGGCACAGAC	483
Qy	104	LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys	123
Db	484	CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAATTTGGAGCCTCTCGAAACCCCTGT	540
Qy	124	SerGluThrTyrCysGlyLeuTyrProGlnSerGluProGluValLysAlaValAlaSer	143
Db	541	GATGAACACTTACTGTGGACCTGCCCGACAGTCTGAAAGGAGACCAAGGCCCTGGCTGAT	600
Qy	144	PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln	163
Db	601	TTATCCGCGCAAAACTCTCTCCATCAAGGCATATCTGACATNTCCACTGTGTACTCCAA	660
Qy	164	HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer	183
Db	661	ATGATGATCTACCCCTTACTCATATGCTTTACAAACTCGGTGAGAAACAATGCTGAGTTGAAT	720
Qy	184	LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr	203
Db	721	GCCTGGCTAAAGCTACTGTGAAGAAGACTT---GCCTCAGTCGACGGCACCAAGTACACA	777
Qy	204	HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr	223
Db	778	TATGGCCGGGAGCTACAAACAATCTATCTCTGCTGGGGCTCTGACGAGTGGCTTAT	837
Qy	224	AspLeuGlyIleLysTyrSerPheThr	232
Db	838	GACCAAGGAATCAGATATTCTTTACC	864

RESULT 13

US-09-011-769A-46

US 03 011 038 A
; Sequence 46, Application US/09011769A

```

; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011,769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1047
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 67..1047
; . SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-011-769A-46

Alignment Scores:
Pred. No.: 6,93e-62 Length: 11
Score: 567.00 Matches: 11
Percent Similarity: 58.36% Conservative: 4
Best Local Similarity: 41.26% Mismatches: 7
Query Match: 42.38% Indels: 4
DB: 4 Gaps: 5

US-09-980-881A-4 (1-246) x US-09-011-769A-46 (1-1053)

Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGln
|||...|||...|||...|||...|||...
Db 67 GCACCTGGTCACTCTTACGAGAGTACACAAAGTGGGAAC

Qy 21 PheIleThrGluArgHisProAspMetLeuThrIysIleHis
|||...|||...|||...|||...|||...
Db 127 CAAGTCGCCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGT

Qy 41 LysTyrProLeuTyrValLeuIysValSerGlyIysGluGln
|||...|||...|||...|||...|||...
Db 187 GGACGCGGTATTATCACTCTCTCAAGGTT---GGCAAAAGCTGG

```

```
Qy 61 TrrleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 244 TTCATGACTGTGGTTCCATGCCAGAGAGTGGATTCTCTGCAATCTGCCAGTGGTTT 303
Qy 81 Ile-----81
Db 304 GTAAGAGAGGCTGTTGCTACTATGAGCGTGGAGATCCAAGTGACAGAGCTTCTCGACAAG 363
Qy 82 -----GlyHis-----83
Db 364 TTAGACTTTTATGTCCTGCTGCTGCTCAATATTGATGGCTTACATCTACACCTGGACCAAG 423
Qy 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 103
Db 424 AGCGATTTTGGAGAAAGACTGCTCCACCCTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 483
Qy 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 484 CCCAACAGAAATTT---GATGCTGGTGGTGGTGGAAATTGGAGCTCTCGAAACCCCTGT 540
Qy 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 541 GATGAAATTTACTGTGACCTGCGCAGAGTCTGAAAGAGGAGACCAAGGCGCTGGCTGAT 600
Qy 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 601 TTATCCGCAACAACCTCTCTTCATCAGGCATCTGCAATCCACTCGTACTCCAA 660
Qy 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 661 ATGATGATCTACCTTACTCATATGCTTACAACTCGTGAGAACAAATGCTGAGTTGAAT 720
Qy 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 721 GCCTGCTAAAGCTACTGTGAAGAAGACTT---GCCTCACTGACGCGCACCAAGTACACA 777
Qy 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
Db 778 TATGGCCCGGAGGACTACAAACATCTATCTGCTGGGGGCTCTGACGACTGGGCTTAT 837
Qy 224 AspLeuGlyIleLysTyrSerPheThr 232
Db 838 GACCAAGGAATCAGATATCTTCCACC 864

RESULT 14
US-08-860-882A-56
; Sequence 56, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGLESE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,882A
FILING DATE: JUNE 23, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: DONALD J. BIRD
REGISTRATION NUMBER: 25,323
REFERENCE/DOCKET NUMBER: 9901/238653
TELEPHONE: (202) 861-3027
TELEFAX: (202) 822-0944
TELEX: 6174627 CUSH
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 1263 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-860-882A-56

Alignment Scores:
Pred. No.: 9,31e-62 Length: 1263
Score: 567.00 Matches: 111
Percent Similarity: 58.36% Conservative: 46
Best Local Similarity: 41.26% Mismatches: 72
Query Match: 42.38% Indels: 40
DB: 2 Gaps: 5

US-09-980-881A-4 (1-246) x US-08-860-882A-56 (1-1263)
Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 325 GCACAGGACACAGATTGATGAGAGTACAAAGTGGGAAACGATGAGGCTTGACTCAA 384
Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 385 CAAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTATTCGGAACACCATTTGAG 444
Qy 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 445 GCACGCGCTATTACCTCTCTGAAGTT---GGCAAGAGCTGGACAAATAAGCCTGCCATT 501
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 502 TTCATGACTGTGGTTTCCATGCCAGAGTGGATTCTCTGCAATCTGCCAGTGGTTT 561
Qy 81 Ile-----81
Db 562 GTAAGAGAGGCTGTTGCTACTATGAGCGTGGAGATCCAAGTGACAGAGCTTCTCGACAAG 621
Qy 82 -----GlyHis-----83
Db 622 TTAGACTTTTATGCTGCTGCTGCTCAATATTGATGGCTTACATCTACACCTGGACCAAG 681
Qy 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAsp 103
Db 682 AGCCGATTTTGGAGAAAGACTGCTGCCACCCATCTAGTGTAGTGTAGTGTAGTGTAGT 741
Qy 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 742 CCCAACAGAAATTT---GATGCTGGTGGTGGTGGAAATTGGAGCTCTCGAAACCCCTGT 798
Qy 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 799 GATGAAATTTACTGTGGACTGCGCGCAGAGTCTGAAAGAGGAGACCAAGGCGCTGGCTGAT 858
Qy 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 859 TTCATCCGCAACAACCTCTCTTCATCAAGGCAATATCTGACAAATCCACTCGTACTCCCAA 918
Qy 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
```

Db 919 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAGAACAACTGCTGAGTTGAAT 978
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 979 GCCTCGCTAAAGCTACTGTGAAGAACTT---GCCTCACTGCGCAGCGCACCAAGTACACA 1035
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTyrPheTyr 223
Db 1036 TATGGCCCGGAGCTACAACTATCTCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 1095
QY 224 AspLeuGlyIleLysTyrSerPheThr 232
Db 1096 GACCAAGGAATCAGATATTCCTTACC 1122

RESULT 15

US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 Mb disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/011,769A
FILING DATE: 13-Feb-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB96/01975
FILING DATE: 13-AUG-1996
APPLICATION NUMBER: GB 9612295.7
FILING DATE: 12-JUN-1996
APPLICATION NUMBER: GB 9611019.2
FILING DATE: 25-MAY-1996
APPLICATION NUMBER: GB 9516810.0
FILING DATE: 16-AUG-1995

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 1263 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

SEQUENCE DESCRIPTION: SEQ ID NO: 38:

US-09-011-769A-38

Alignment Scores:

Pred. No.:	9,31e-62	Length:	1263
Score:	567.00	Matches:	111
Percent Similarity:	58.36%	Conservative:	46
Best Local Similarity:	41.26%	Mismatches:	72
Query Match:	42.38%	Indels:	40
DB:	4	Gaps:	5

US-09-980-881A-4 (1-246) x US-09-011-769A-38 (1-1263)

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPheGlu 20
Db 325 GCAACAGGACACAGTATGAGAGTACAAACAAAGTGGAAACGATAGAGGCTTGGACTCAA 384
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 385 CAAGTCGCCACTCAGAGTCCAGCCCTCATCTCTCGAGTGTATTCGGAACACATTGAG 444
QY 41 LysTyrProLeuTyrValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 445 GGACGCGCTATTACCTCTGAAAGTT---GGCAAGCTGGACAAATAAGCCCTGCATT 501
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTyrPhe 80
Db 502 TTCATGGACTGTGTTCATGCCAGAGTGGATTCTCTGCATTCTGCCAGTGGTTT 561
QY 81 Ile----- 81
Db 562 GTAAGAGAGGCTGTTCTGATCTATGACGCTGAGATCCAAGTGACAGAGCTTCTCGACAAG 621
QY 82 -----GlyHis----- 83
Db 622 TTAGACTTTTATGCTCTGCTGCTGCTCAATATTGATGGCTACATCTACACTGGACCAAG 681
QY 84 AsnArgMetTyrArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 682 AGCCGATTTTGGAGAAAGACTCGCTCCACCCACTACTGGATCTAGCTGCATTGGCAGAC 741
QY 104 LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerCys 123
Db 742 CCCAACAGAAATTTT---GATGCTGTTGTGTGAATTTGGAGCCCTCTCGAAACCCCTGT 798
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaLysSer 143
Db 799 GATGAAACTTACTGTGACCTGCGCAGAGTCTGAAAGGAAACCAAGGCCCTGGCTGAT 858
QY 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyrIleSerMethHisSerTyrSerGln 163
Db 859 TTCATCCGCAACAAACTCTTTCATCAAGGCATATCTGACAACTCCACTCGTACTCCCAA 918
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 919 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAGAACAAATGCTGAGTTGAAT 978
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 979 GCCTCGCTAAAGCTACTGTGAAGAACTT---GCCTCACTGCGCAGCGCACCAAGTACACA 1035
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTyrPheTyr 223
Db 1036 TATGGCCCGGAGCTACAACTATCTATCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 1095
QY 224 AspLeuGlyIleLysTyrSerPheThr 232
Db 1096 GACCAAGGAATCAGATATTCCTTACC 1122

Search completed: January 2, 2004, 09:26:07

Job time : 125.985 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 20:10:07 ; Search time 598.309 Seconds
(without alignments)
1423.134 Million cells updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQVHSLNEIYSWIE.....IKVSTSNPPVKKLPLSLK 246

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cg2_1/USPTO_spool/US0980881/runat_31122003_135919_26949/app_query.fasta_1.1628
-DB=Published Applications NA -QWTF=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0980881 @CGN 1 1 704 @runat_31122003_135919_26949
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cg2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cg2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cg2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cg2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cg2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cg2_6/ptodata/1/pubpna/US09_PUB.seq:
14: /cg2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cg2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17: /cg2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18: /cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1272	95.1	1728	10	US-09-880-107-2396	Sequence 2396, Ap
2	1177	88.0	1272	13	US-10-379-836-1	Sequence 1, Appli
3	902	67.4	1400	9	US-09-925-302-24	Sequence 24, Appli
4	556.5	41.6	1311	14	US-10-200-344-9	Sequence 9, Appli
5	556.5	41.6	1311	14	US-10-200-344-9	Sequence 33, Appli
6	552	41.3	1254	15	US-10-229-546-3	Sequence 3, Appli
7	552	41.3	1254	15	US-10-229-546-3	Sequence 21, Appli
8	552	41.3	1622	12	US-10-429-802-21	Sequence 1, Appli
9	552	41.3	1622	12	US-10-429-802-21	Sequence 187, App
10	550.5	41.1	1125	9	US-09-888-615-2	Sequence 2, Appli
11	550.5	41.1	1332	16	US-10-176-306-75	Sequence 75, Appli
12	550.5	41.1	1603	16	US-10-176-306-75	Sequence 73, Appli
13	549	41.0	1870	10	US-09-910-059-112	Sequence 112, App
14	549	41.0	1870	10	US-09-910-059-112	Sequence 124, App
15	547.5	40.9	1332	15	US-09-954-456-1141	Sequence 1141, Ap
16	543.5	40.6	1826	13	US-10-252-157-453	Sequence 453, App
17	525.5	39.3	2128	14	US-10-200-344-13	Sequence 13, Appli
18	492	36.8	1200	15	US-10-200-910-7	Sequence 7, Appli
19	471.5	35.2	1050	14	US-10-200-344-11	Sequence 11, Appli
20	465.5	34.8	1254	9	US-09-925-297-58	Sequence 58, Appli
21	465.5	34.8	1260	13	US-10-345-680-66	Sequence 66, Appli
22	465.5	34.8	1311	15	US-10-200-910-5	Sequence 5, Appli
23	465.5	34.8	1380	10	US-09-969-347-209	Sequence 209, App
24	465.5	34.8	1380	13	US-10-345-680-64	Sequence 64, Appli
25	458	34.2	2786	11	US-09-946-374-233	Sequence 233, App
26	458	34.2	2786	13	US-10-015-387A-233	Sequence 233, App
27	458	34.2	2786	13	US-10-006-130A-233	Sequence 233, App
28	458	34.2	2786	13	US-10-199-672-307	Sequence 307, App
29	458	34.2	2786	13	US-10-006-172A-233	Sequence 233, App
30	458	34.2	2786	13	US-10-187-749-307	Sequence 307, App
31	458	34.2	2786	13	US-10-194-457-307	Sequence 307, App
32	458	34.2	2786	13	US-10-184-642-307	Sequence 307, App
33	458	34.2	2786	13	US-10-196-747-307	Sequence 307, App
34	458	34.2	2786	13	US-10-015-392A-233	Sequence 233, App
35	458	34.2	2786	13	US-10-017-253A-233	Sequence 233, App
36	458	34.2	2786	13	US-10-173-690-307	Sequence 307, App
37	458	34.2	2786	13	US-10-173-691-307	Sequence 307, App
38	458	34.2	2786	13	US-10-173-692-307	Sequence 307, App
39	458	34.2	2786	13	US-10-173-694-307	Sequence 307, App
40	458	34.2	2786	13	US-10-173-698-307	Sequence 307, App
41	458	34.2	2786	13	US-10-173-698-307	Sequence 307, App
42	458	34.2	2786	13	US-10-173-699-307	Sequence 307, App
43	458	34.2	2786	13	US-10-173-707-307	Sequence 307, App
44	458	34.2	2786	13	US-10-174-569-307	Sequence 307, App
45	458	34.2	2786	13	US-10-174-583-307	Sequence 307, App

ALIGNMENTS

RESULT 1
US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Alignment Scores:
Pred. No.: 3,06e-160 Length: 1728
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-4 (1-246) x US-09-880-107-2396 (1-1728)

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
DB 362 GCCTCCGCATCGTACTATGACAGTATCACTCAATAATGAATCTATCTTGGATAGAA 421

QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 422 TTTATAACTGAGAGGCATCTGATATGCTTACAAAAATCCACATTTGGATCCTCATTTGAG 481

QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 482 AAGTACCACCTCTATGTTTTAAAGGTTCTGGAAAAAGAACAAACAGCAAAAATGCCATA 541

QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrTrpIleSerProAlaPheCysLeuTyrPhe 80
DB 542 TGGATTGACTGTGGATCCATGTCAGAGATGGATCTCTCTGCTTTCTGCTTGTGGTTC 601

QY 81 IleGlyHis----- 83
DB 602 ATAGGCCATATACTCAATTTCTATGGGATAATAGGCAATATACCAATCTCTGAGGCTT 661

QY 83 ----- 83
DB 662 GTGGATTTCTATGTTATGCGGTGTTAATGTGGACGGTTATGACTACTCATGCGAAAAAG 721

QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
DB 722 AATCGAATGTGGAAAGAACCGTTCTTTCTATGCGAACCAATCATTTGCATCGGAACAGAC 781

QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
DB 782 CTGAATAGGAACCTTGCTTCCAAACACTGTGTGAGGAAGTGCATCCAGTTCTCTATGC 841

QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
DB 842 TCGGAAACCTTACTGTGACTTTATCTCTGAGTCAGAACCAAGTGAAGGCGAGTGGCTAGT 901

QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
DB 902 TTTCTTGAGAGAATAATATCAACAGATTAAGCAATACATCAGCATGCAATCATCTCCAG 961

QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
DB 962 CATATAGTGTTCATATCTCTATACAGAAAGTAAAGCAACCAATGAGGAACTGTCT 1021

QY 184 LeuValaSerGluAlaValaAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
DB 1022 CTAGTAGCCAGTGAAGCAGTTCTGCTATTGAGAAAACTAGTAAAAATACAGGTATACA 1081

QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
DB 1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGACGATGGATCTAT 1141

QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
DB 1142 GATTTGGGCAATCAATATTCGTTTACAACTTCGAGATAGGGCAATACGAGATTC 1201

QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
DB 1202 TTGCTGCGGAGCGTTACATCAAAACCCACCTGTAGAGAACTTTTGGCGGTCTCTTAAA 1261

QY 246 s 246
DB 1262 A 1262

RESULT 2

US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Papio hamadryas
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1269)
US-10-379-836-1

Alignment Scores:
Pred. No.: 1.13e-147 Length: 1272
Score: 1177.00 Matches: 228
Percent Similarity: 78.41% Conservative: 8
Best Local Similarity: 75.75% Mismatches: 10
Query Match: 87.97% Indels: 55
DB: 13 Gaps: 2

US-09-980-881A-4 (1-246) x US-10-379-836-1 (1-1272)

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
DB 343 GCCTCCGCATCGTACTATGACAGTATCACTCAATAATGAATCTATCTTGGATAGAA 402

QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 403 CTTATTAAGTGAAGATGATCTCTGATATGCTTCAAAAATCCACATTTGGATCCTCTATGAG 462

QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 463 AAGCACCACCTTTATGTTTTAAAGGTTTCTGGAAAAAGAACAAACAGCAAAAATGCCATG 522

QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrTrpIleSerProAlaPheCysLeuTyrPhe 80
DB 523 TGGATTGACTGTGGAATCCATGCCAGAAATGGATCTCCCTCTGCTTTCTGCTTGTGGTTC 582

QY 81 IleGlyHis----- 83
DB 583 ATAGGCCATATACTGAATACTACGGGATATAGGGGAATATACCAATCTTCTGAGGCAT 642

QY 83 ----- 83
DB 643 GTGGATTTCTATCTATGCCAGTGGTTAATGTGGATGGTTTATGACTACTCATGAAAAAG 702

QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
DB 703 AATCGAATGTGGAAAAAGAACCGTTCTTTCTATCGCAACAATCGTTGTCATCGGAACAGAC 762

QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
DB 763 CTGAACAGGAACCTTTCGTCCTCAAAACACTGGTGTGGAGGAAGTGCATTCAGTTCTCTAAGC 822

QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
DB 823 TCGGAAACCTTACTGTGGACTTTATCTGAGTCAAGACCAAGAGGAGCGGTGGCTAAT 882

QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
DB 883 TTCTTGAGAGAAATATCAACCACTATAAGACATACATCAGCATGCACTACTCCAG 942
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
DB 943 CATATCGTGTTCATATTCCTATATCTCGNAGCAAAAGCAACACAGGAAATGTCT 1002
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
DB 1003 CTAGTAGCCAGTGAAGCAGTTCTGCTATTTCAGAAACCACTATAATATCAGGTATACA 1062
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyClyAspThrIleTyr 223
DB 1063 CATGGCCGTGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGCGACGATTGATCTAT 1122
QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
DB 1123 GATTGGGCATCAAAATATTCGTTTACAATGTGAACCTTCGAGATACGGGCAATACGGATTTC 1182
QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
DB 1183 TTGCTGCTGAGGTTTACATCAAAACCCACTTGTGAAGAGCGCTTTTGGCGTGTCTCTAAA 1242
QY 246 S 246
DB 1243 A 1243

RESULT 3

US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (25)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Alignment Scores:
Pred. No.: 1.36e-110 Length: 1400
Score: 902.00 Matches: 181
Percent Similarity: 76.05% Conservative: 0
Best Local Similarity: 76.05% Mismatches: 1
Query Match: 67.41% Indels: 56
DB: 9 Gaps: 2

US-09-980-881A-4 (1-246) x US-09-925-302-24 (1-1400)

QY 65 GlyIleHisAlaArgGluThrIleSerProAlaPheCysLeuTrpPheLeuHis---- 83
DB 10 GGAATCCATGCCAGANAATGGATCTCTCTGCTTCTGCTTGTGGTTTCATAGGCCATA 69
QY 83 ----- 83
DB 70 ACTCAATTTCTATGGGTAATAGGGCAATATACCAATCTCTGAGGCTTGTGGATTTCTAT 129

QY 84 -----AsnArgMetTr 87
DB 130 GTTATGCCGGTGGTTAATGTGGATGTTATGAACCTACTCATGCAAAAGAAATCGAATGTG 189
QY 87 PArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAs 107
DB 190 GAGNAGAAGACCGTCTTCTTCTATGCGAACAATCATTCGATCGGAACAGACCTGANTAGGAA 249
QY 107 nPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerSerCysSerGluThrTy 127
DB 250 CTTTGTCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309
QY 127 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 147
DB 310 CTGTGACATTTATCTCTGAGTCAGAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 369
QY 147 GAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 167
DB 370 AAATATCAACCGAGATTAAGCATATCATCAGCATGCATTCATCTCCAGCATATAGTGT 429
QY 167 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 187
DB 430 TCCATATTCCTATACAGGAAGTAAAGCAAGACCATGAGGAACCTGCTCTAGTAGCCAG 489
QY 187 rGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 207
DB 490 TGAAGCAGTTCGTGCTATTGAGAAAACCTAGTAAAAATACCAAGGTATACATGCGCCATGG 549
QY 207 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTyrIleTyrAspLeuGlyIl 227
DB 550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGG 609
QY 227 elyTyrSerPhe----- 231
DB 610 CAAATATTCGTTTACAATTTGAACTTCGAGATACGGGCACATACGGATTCTTGTCTGCCGA 669
QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
DB 670 GCGTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCGCTGTCTCTAAAA 719

RESULT 4

US-10-200-344-9
; Sequence 9, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-200-344-9

Alignment Scores:
Pred. No.: 3.86e-64 Length: 1311
Score: 556.50 Matches: 113
Percent Similarity: 53.45% Conservative: 42
Best Local Similarity: 38.97% Mismatches: 77
Query Match: 41.59% Indels: 58
DB: 14 Gaps: 5

```
US-09-980-881A-4 (1-246) x US-10-200-344-9 (1-1311)
Qy 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
Db 406 TATGAAGTTTATCACTCTCTAGAGAAATTCAAATTTGGATGCATCATCTGAATAAACT 465
Qy 26 HisProAspMetLeuThrLysIleHisIleGlySerPheGluLysTyrProLeuTyr 45
Db 466 CACTCAGGCTTCATCATGTTCTCTATTGGAGATCATATGAGGGAAGATCTCTTTT 525
Qy 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
Db 526 ATTTTAAAGCTG---GGCAGCAGCATCAGCACTCAAAGAGCTGTTTGGATAGACTGGT 582
Qy 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle 81
Db 583 ATTCATCAAGAGAATGGATTGGTCTCTGCTCTTTTGTAGTGGTTTGTAAAGAAGCTCTT 642
Qy 81 ----- 81
Db 643 CTAACATATAAGAGTACCAGCCATGAGAAAATTTGAATCATCTATATTTCTATATC 702
Qy 82 -----GlyHisAsnArgMetTrpArg 88
Db 703 ATGCTGTGTTTAACGTCGATGATACCAATTTAGTTGGACCAATGATCGATTGGAGA 762
Qy 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
Db 763 AAAACAAGGTCAGGAACCTCAAGGTTTCGCTGCGTGGAGTGCATGCAATAGAAAATGG 822
Qy 109 AlaSerLysHisTrpCysGluGlnGlyValAsnSerSerSerCysSerGluThrTyrCys 128
Db 823 AAGCTGAAG---TGTGTGATGAAGAGAGCTTCTATGCACCTTGTGTGATGACATACGT 879
Qy 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
Db 880 GGCCCTTTTCCAGATCTGAGCGGAAGTGAAGGCTGTAGTAACTTCTTCGAAACAC 939
Qy 149 IleAsnGlnLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
Db 940 AGAAAGCACATTAGGGCTTATCTCTCTTTCATGATATGCTCAGATGTTACTGTATCCC 999
Qy 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
Db 1000 TATTCCTTACAAATATGCAACAATTCCTCAATTTAGATGTGTGGAATCTGCAGCTTATAA 1059
Qy 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
Db 1060 GCTGTGAATGCATCTCAGTCAGTA---TACGGGGTACAGATACAGATATGACCGCTCC 1116
Qy 209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGlyIleLys 228
Db 1117 ACAACGTTGTATGTAGCTCTGCTGATGCTCAATGATGGTGGCTACAAAATGGAATACCT 1176
Qy 229 TyrSerPhe----- 231
Db 1177 TATGCATTTGCTTTCGAACCTACGTGACCTGGATATTTTGGATTTTCTCCAGAGATG 1236
Qy 232 -ThrSerAsnProProValGluLysLeu 240
Db 1237 CTCATCAAAACCCACTGTACAGAACTA 1264

RESULT 5
US-10-274-639-33
; Sequence 33, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DELEGANE, Angelo M.; GANDHI, Ameena R.
; APPLICANT: HAFALIA, April J.A.; LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
```

```

; APPLICANT: NGUYEN, Damiel B.; LEE, Ernestine A.
; APPLICANT: KHAN, Farrah A.; YUE, Henry
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha
; APPLICANT: DING, Li; KEARNEY, Liam
; APPLICANT: SAUNJANWALA, Madhusudan M.; YAO, Monique G.
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.
; APPLICANT: LAL, Preeti G.; LEE, Sally
; APPLICANT: TODD, Stephen; LO, Terence P.
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
; APPLICANT: AZIMZAI, Valda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,717
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/225,988
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 60/227,568
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 33
; LENGTH: 1993
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CB1
; US-10-274-639-33

Alignment Scores:
Pred. No.: 7 68e-64 Length: 1993
Score: 556.50 Matches: 113
Percent Similarity: 53.45% Conservative: 42
Best Local Similarity: 38.97% Mismatches: 77
Query Match: 41.59% Indels: 58
DB: 12 Gaps: 5

US-09-980-881A-4 (1-246) x US-10-274-639-33 (1-1993)
Qy 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
Db 743 TATGAAGTTTATCACTCTCTAGAGAAATTCAAATTTGGATGCATCATCTGAATAAACT 802
Qy 26 HisProAspMetLeuThrLysIleHisIleGlySerPheGluLysTyrProLeuTyr 45
Db 803 CACTCAGGCTTCATCATGTTCTCTATTGGAGATCATATGAGGGAAGATCTCTTTT 862
Qy 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
Db 863 ATTTTAAAGCTG---GGCAGCAGCATCAGCACTCAAAGAGCTGTTTGGATAGACTGGT 919
Qy 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle 81
Db 920 ATTCATCAAGAGAATGGATTGGTCTCTGCTCTTTTGTAGTGGTTTGTAAAGAAGCTCTT 979
Qy 81 ----- 81
Db 980 CTAACATATAAGAGTACCAGCCATGAGAAAATTTGAATCATCTATATTTCTATATC 1039
Qy 82 -----GlyHisAsnArgMetTrpArg 88
```

1040 ATGCCTGTGTTAACTGCGATGATACCATTTTGTGGACCAATGATCGATTTTGGAGA 1099
QY 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
Db 1100 AAAACAAGGTCAAGGAACCTCAAGGTTTCGTCGCGTGGAGTGATGCCAATAGAAACTGG 1159
QY 109 AlaSerLysHisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 128
Db 1160 AAGTGAAG---TGGTGTGATGAAGAGCTTCTATGCACCCCTTGTGATGACACATAGTGT 1216
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
Db 1217 GGGCTTTTCCAGAACTGAGCCGGAAGTGAAGGCTGTAGCTTAACCTTCTTGGAAAACAC 1276
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
Db 1277 AGAAAGCACATAGGGCTTATCTCTCTTTCATGCATATGCTCAGATGTTACTGTATCCC 1336
QY 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
Db 1337 TATCTTCAAAATATCAACAATTCCTCAATTTTGTAGATGTGTGGAATCTGCAGCTTATAAA 1396
QY 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
Db 1397 GCTGTGAATGCACTTCAGTCAGTA---TACGGGTAGCATACAGATATGAGCCAGCTCC 1453
QY 209 GluThrLeuTyrLeuAlaProGlyGlyAspTyrIleTyrAspLeuGlyIleLys 228
Db 1454 ACAACGTTGTATGTAGCTCTGCTAGCTCAATGATGGGCTACAAAATGGAATACCT 1513
QY 229 TyrSerPhe----- 231
Db 1514 TATGCATTTGCTTTCGAACCTACGTGACACTGGATATTTTGGATTTTACTCCAGAGATG 1573
QY 232 -ThrSerAsnProProValGluLysLeu 240
Db 1574 CTATCAAAACCCACCTGTACAGAAACTA 1601

RESULT 6

US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; FILE REFERENCE: MPI01-156PIRNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; PRIOR FILING DATE: 2002-08-28
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-229-546-3
Alignment Scores:
Pred. No.: 1,45e-63 Length: 1254
Score: 552.00 Matches: 109
Percent Similarity: 58.56% Conservative: 45
Best Local Similarity: 41.44% Mismatches: 69
Query Match: 41.26% Indels: 40
DB: 15 Gaps: 5

RESULT 7

US-10-429-802-21
; Sequence 21, Application US/10429802
; Publication No. US2003028285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54

US-09-980-881A-4 (1-246) x US-10-229-546-3 (1-1254)
QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrIleGluPheIleThrGluArg 25
Db 346 TAGCGAAAATACATAAATATGGGAAAAGATTGGCTTGGACTGAAAGATGATGATAAG 405
QY 26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
Db 406 TATCTCGAATGGTCTCTCGTATTAAATGGATCTACTGTTGAAGATTAATCCACTATAT 465
QY 46 ValLeuValSerGlyLysGluGlnThrAlaLysAsnAlaIleTyrIleAspCysGly 65
Db 466 GTTCTGAAGATT---GGGAAAAGATGAAGAGAAAGCTATTTTATGGATGCTGGC 522
QY 66 IleHisAlaArgGluTyrIleSerProAlaPheCysLeuTyrPheIle----- 81
Db 523 ATTACGCGACGAGAATGGGTCTCCCGAGCATTTCTGCAGTGGTGTGCTCTATCAGGCAACC 582
QY 81 ----- 81
Db 583 AAAACTTATGGGAGAAACAAAATATGACCAAACTCTTGGACCGAATGAATTTTACATT 642
QY 82 -----GlyHis-----AsnArgMetTyrArg 88
Db 643 CTTCTGTGTTCAATGTTGATGATATATTTGGTCATGACAAAGAACCGCATGTGAGA 702
QY 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
Db 703 AAAATCGTTTCAAGAACCAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT 762
QY 109 AlaSerLysHisTyrCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 128
Db 763 ---AATGCTTCATGGAACCTCCATTCTTAACCAATGACCCCATGTGCAGATAACTATCGG 819
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
Db 820 GGCTCTGCACAGAGTCCGAGAAAGAGACGAAGCTGCTCACTAATTTCTATTAGAACCCAC 879
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
Db 880 CTGAATGAATCAAGGTTTACATCACCTTCCATTCTCTACTCCAGATGCTATTGTTTCCC 939
QY 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
Db 940 TATGATATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCCAAAGATT 999
QY 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
Db 1000 GGCACTGATGTTCTA---TCAACTCGATATGAACCCGCTACATCTATGGCCCAATAGAA 1056
QY 209 GluThrLeuTyrLeuAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLys 228
Db 1057 TCAACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAA 1116
QY 229 TyrSerPhe 231
Db 1117 CACACATTT 1125

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 21

; LENGTH: 1622

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-10-429-802-21

Alignment Scores:

Pred. No.: 2,21e-63 Length: 1622
Score: 552.00 Matches: 109
Percent Similarity: 58.56% Conservative: 45
Best Local Similarity: 41.44% Mismatches: 69
Query Match: 41.26% Indels: 40
DB: 5 Gaps: 5

US-09-980-881A-4 (1-246) x US-10-429-802-21 (1-1622)

```
QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
DB 346 TACGAAATACATAATTCGGAAGAGATTGGCTTGGACTGAAAGATGATGGAATAG 405
QY 26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
DB 406 TATCCTGAAATGGTCTCTCGTATTAATAATTGGATCTACTGTGAAGATAATCCACTATAT 465
QY 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
DB 466 GTTCTGAAGATT---GGGGAAGAAGATGAAGAAGAGAGGCTATTTTATGGATTGGGC 522
QY 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 81
DB 523 ATTCACGCACGAGATGGGTCTCCCGACGATCTCGCCAGTGGTTTGTCTATCAGGCAACC 582
QY 81 ----- 81
DB 583 AAAAATTATGGGAGAAACAAAATTATGACCAAACTCTTGACCGCAATGAATTTTACATT 642
QY 82 -----GlyHis-----AsnArgMetTrpArg 88
DB 643 CTTCCTGTGTTCATGTGGATATATTTGGTCTGGAACAAGACCCGATGTGGAGA 702
QY 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
DB 703 AAAAATCGTTCCAAAGAACCAAACTCCAATGCGACTGACCTCAACAGAAATTTT 762
QY 109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerSerSerSerSerSerSer 128
DB 763 ---AATGCTTCATGGAATCCATTCCTACACCAATGACCCATGTGCAGATACTATCG 819
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
DB 820 GGCTCTGCACGAGTCCGAGAACGAGAAAGAGACGAGTGTCTATGAGCTGGGCAATCAA 879
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisTyrSerGlnHisIleValPhePro 168
DB 880 CTGAATGAATCAAGTTTACATCACCTTCAATCTCTACTCCGATGCTATTTGTTCCC 939
QY 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuAlaLysGlu 188
DB 940 TATGGATATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAGATT 999
QY 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
DB 1000 GGCACTGATGTTCTA-----TCAACTCGATATGAACCCGCTACATCTATGTCGCAATAGAA 1056
QY 209 GluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
DB 1057 TCAACATTTACCGATATCAGTTCTTCTTAGACTGGGCTTATGACCTGGGCAATCAA 1116
QY 229 TyrSerPhe 231
DB 1117 CACACATTT 1125
```

RESULT 8

US-10-229-546-1

; Sequence 1, Application US/10229546

; Publication No. US20030082649A1

; GENERAL INFORMATION:

; APPLICANT: Weich, Nadine S.

; APPLICANT: Carroll, Joseph M.

; APPLICANT: Millennium Pharmaceuticals, Inc.

; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE

; FILE REFERENCE: FAMILY MEMBER AND USES THEREFOR

; CURRENT FILING DATE: 2002-08-28

; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; TYPE: DNA

; LENGTH: 1622

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1254)

US-10-229-546-1

Alignment Scores:

Pred. No.: 2,21e-63 Length: 1622
Score: 552.00 Matches: 109
Percent Similarity: 58.56% Conservative: 45
Best Local Similarity: 41.44% Mismatches: 69
Query Match: 41.26% Indels: 40
DB: 5 Gaps: 5

US-09-980-881A-4 (1-246) x US-10-229-546-1 (1-1622)

```
QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
DB 346 TACGAAATACATAATTCGGAAGAGATTGGCTTGGACTGAAAGATGATGGAATAG 405
QY 26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
DB 406 TATCCTGAAATGGTCTCTCGTATTAATAATTGGATCTACTGTGAAGATAATCCACTATAT 465
QY 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
DB 466 GTTCTGAAGATT---GGGGAAGAAGATGAAGAAGAGGCTATTTTATGGATTGGGC 522
QY 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----- 81
DB 523 ATTCACGCACGAGATGGGTCTCCCGACGATCTCGCCAGTGGTTTGTCTATCAGGCAACC 582
QY 81 ----- 81
DB 583 AAAAATTATGGGAGAAACAAAATTATGACCAAACTCTTGACCGCAATGAATTTTACATT 642
QY 82 -----GlyHis-----AsnArgMetTrpArg 88
DB 643 CTTCCTGTGTTCATGTGGATATATTTGGTCTGGAACAAGACCCGATGTGGAGA 702
QY 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
DB 703 AAAAATCGTTCCAAAGAACCAAACTCCAATGCGACTGACCTCAACAGAAATTTT 762
QY 109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerSerSerSerSerSerSer 128
DB 763 ---AATGCTTCATGGAATCCATTCCTACACCAATGACCCATGTGCAGATACTATCG 819
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
DB 820 GGCTCTGCACGAGTCCGAGAACGAGAAAGAGACGAGTGTCTATGAGCTGGGCAATCAA 879
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisTyrSerGlnHisIleValPhePro 168
```

```
Db 880 CTGAATGAATCAAGTTTACATCACTTCCATCTTCTACCTCCAGATGCTATGTTTCC 939
      ::::::::::::::::::::|
Qy 169 TyrSerTyrThrArgSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
      ::::::::::::::::::::|
Db 940 TATGGATATACATCAAACTGCCACCTAACCATGAGGACTTGCCCAAGTTGCAAGATT 999
      ::::::::::::::::::::|
Qy 189 AlavalargAlalleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
      ::::::::::::::::::::|
Db 1000 GGCACCTGATGTTCTTA---TCAACTCGATATGAACCCGCTACATCTATGCGCCCAATAGAA 1056
      ::::::::::::::::::::|
Qy 209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTyrPileTyrAspLeuGlyLys 228
      ::::::::::::::::::::|
Db 1057 TCAACCAATTTACCGATATCAGGTTCTTCTTGAAGCTGGCTTATGACCTGGGCATCAAA 1116
      ::::::::::::::::::::|
Qy 229 TyrSerPhe 231
      ::::::::::::::::::::|
Db 1117 CACACATTT 1125
      ::::::::::::::::::::|

RESULT 9
US-10-341-434-187
; Sequence 187, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
; LENGTH: 1633
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)..(1262)
; OTHER INFORMATION:
US-10-341-434-187

Alignment Scores:
Pred. No.: 2-23e-63 Length: 1633
Score: 552.00 Matches: 109
Percent Similarity: 58.56% Conservative: 45
Best Local Similarity: 41.44% Mismatches: 69
Query Match: 41.26% Indels: 40
DB: 13 Gaps: 5

US-09-980-881A-4 (1-246) x US-10-341-434-187 (1-1633)
Qy 6 TyrGluGlnTyrHisSerLeuAsnGluLeuTyrSerTrpIleGluPheLeuThrGluArg 25
      ::::::::::::::::::::|
Db 357 TACGCAAAATACAAATAATTTGGGAAAGATTGTGCTTGGACTGAAAGATGATGGATAAG 416
      ::::::::::::::::::::|
Qy 26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
      ::::::::::::::::::::|
Db 417 TATCCTGAATGGTCTCTCGTATTAATAATTTGGATCTACTGTTGAAGATAATCCACTATAT 476
      ::::::::::::::::::::|
Qy 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
      ::::::::::::::::::::|
Db 477 GTTCTGAAGATT---GGGAAAGATGAAGAAAGAGCGCTATTTTATGGATTGGC 533
      ::::::::::::::::::::|
Qy 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheLeu----- 81
      ::::::::::::::::::::|
Db 534 ATTCAGCAGAGATGGGTCTCCCGAGCATTTCCGAGGTGTTTGTCTATCAGGCAACC 593
      ::::::::::::::::::::|
Qy 81 ----- 81
```

```
Db 594 AAAACTTATGGGAGAAACAAAATTATGACCAAACTCTTGGACGATGAATTTTACATT 653
      ::::::::::::::::::::|
Qy 82 -----GlyHis-----AsnArgMetTrpArg 89
      ::::::::::::::::::::|
Db 654 CTTCTGTGTTCAATGTTGATGGATATATTTGGTCAATGACAAAGAACCGCATGTGAGA 713
      ::::::::::::::::::::|
Qy 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
      ::::::::::::::::::::|
Db 714 AAAAATCGTTTCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTT 773
      ::::::::::::::::::::|
Qy 109 AlaSerLysHisTyrCysGluGlyAlaSerSerSerCysSerGluThrTrpCys 128
      ::::::::::::::::::::|
Db 774 ---AATGCTTCATGGAACCTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCG 830
      ::::::::::::::::::::|
Qy 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
      ::::::::::::::::::::|
Db 831 GGCCTCTGCACAGAGTCCGAGAAAGAGAGAGAAAGCTGTCAATAATTTCAATAGAGCCAC 890
      ::::::::::::::::::::|
Qy 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
      ::::::::::::::::::::|
Db 891 CTGAATGAATCAAGGTTTACATCACCTTCCATTCCTACTCCAGATGCTATTTGTTTCCC 950
      ::::::::::::::::::::|
Qy 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
      ::::::::::::::::::::|
Db 951 TATGGATATACATCAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAGATT 1010
      ::::::::::::::::::::|
Qy 189 AlavalargAlalleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
      ::::::::::::::::::::|
Db 1011 GGCACCTGATGTTCTTA---TCAACTCGATATGAACCCGCTACATCTATGCGCCCAATAGAA 1067
      ::::::::::::::::::::|
Qy 209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTyrPileTyrAspLeuGlyLys 228
      ::::::::::::::::::::|
Db 1068 TCAACCAATTTACCGGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAA 1127
      ::::::::::::::::::::|
Qy 229 TyrSerPhe 231
      ::::::::::::::::::::|
Db 1128 CACACATTT 1136
      ::::::::::::::::::::|

RESULT 10
US-09-888-615-2
; Sequence 2, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARVDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDAKSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1125
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-888-615-2

Alignment Scores:
Pred. No.: 1-93e-63 Length: 1125
Score: 550.50 Matches: 104
Percent Similarity: 56.44% Conservative: 45
Best Local Similarity: 39.39% Mismatches: 76
Query Match: 41.14% Indels: 39
DB: 9 Gaps: 4

US-09-980-881A-4 (1-246) x US-09-888-615-2 (1-1125)
```



```
QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
Db 139 TATAACATATACCACCCATGGGAGAGATCTATGAGTGGATGAGAGATCACTGAGAAG 198
QY 26 HisProAspMetLeuThrLysIleHisGlySerSerPheGluLysTyrProLeuTyr 45
Db 199 TACAAGGAAGTGGTGACACAGCATTTCTTAGGAGTGCATTCATGAGACCCACCATGTAT 258
QY 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
Db 259 TATCTGAAGATCAGCCCAACCATCTGTAATCCCAAGAAATCAATTTGGATGGACTGTGA 318
QY 66 IleHisAlaArgLutrpIleSerProAlaPheCysLeuTrpPheIle----- 81
Db 319 ATTCAGCCAGAGATGGATTGCTCTGCTCTTTTGCCAATGGTTCGTCAAGAAATCTTA 378
QY 81 ----- 81
Db 379 CAAAACCATAAAGAACACTCAAGTATACGCAAGCTCCTTAGGAACCTGGACTTCATGTC 438
QY 82 -----GlyHis-----AsnArgMetTrpArg 88
Db 439 CTTCACGATCTTAACATAGATGGTTATATATACACTTGGCAACTGATGCTCTTTGAGG 498
QY 89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
Db 499 AATCCGGTTCACCCATAATATGGCACATGTTTGGGACGATCTCAATCGAATTC 558
QY 109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 128
Db 559 ---AATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACTCTGT 615
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
Db 616 GGGACAGCGCCAGTGTCAACACAGACACTAAAGCTGTTCCAGCTTCATAGAGACAAG 675
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
Db 676 AAGGATGATATTGTGCTTCCTGACCATGTCATCTTATGGGAGTTAATTCACACCT 735
QY 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
Db 736 TAGCGGTACACCAAAATAAATCAAGTAACACCCAGAAATGATTCAGAGTTGGACAGAAG 795
QY 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
Db 796 GCAGCAAAATCATG---AAAGCAAAAGTATGGAAACCAATATATAGAGTTGGATCGAGTGA 852
QY 209 GluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
Db 853 GATATTTATATGCTTCATCAGGGTCTTCAAGAGATTGGCCCGAGACATTTGGATTCCC 912
QY 229 TyrSerPheThr 232
Db 913 TTCTCATATACG 924
RESULT 11
US-10-176-306-75
; Sequence 75, Application US/10176306
; Publication NO. US20030130485A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
; FILE REFERENCE: 1048-195001
; CURRENT APPLICATION NUMBER: US/10/176,306
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 10/001,137
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: PCT/US01/45291
```

```
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/248,362
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,331
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,365
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/250,077
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,327
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,176
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 10/023,617
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/49416
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,249
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,405
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 10/083,248
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/US01/46717
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242,324
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,518
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-176-306-75

Alignment Scores:
Pred. No.: 2,556-63 Length: 1332
Score: 550.50 Matches: 104
Percent Similarity: 56.44% Conservative: 45
Best Local Similarity: 39.39% Mismatches: 76
Query Match: 41.14% Indels: 39
DB: 16 Gaps: 4

US-09-980-881A-4 (1-246) x US-10-176-306-75 (1-1332)
QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
Db 529 TATAACATATACCACCCATGGGAGAGATCTATGAGTGGATGAGAGATCACTGAGAAG 588
QY 26 HisProAspMetLeuThrLysIleHisGlySerSerPheGluLysTyrProLeuTyr 45
Db 589 TACAAGGAAGTGGTGACACAGCATTTCTTAGGAGTGCATTCATGAGACCCACCATATAT 648
QY 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
Db 649 TATCTGAAGATCAGCCCAACCATCTGGAATAATCCCAAGAAATCAATTTGGATGGACTGTGA 708
QY 66 IleHisAlaArgLutrpIleSerProAlaPheCysLeuTrpPheIle----- 81
Db 709 ATTCAGCCAGAGATGGATTGCTCTGCTCTTTTGCCAATGGTTCGTCAAGAAATCTTA 768
QY 81 ----- 81
Db 769 CAAAACCATAAAGAACACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCATGTC 828
QY 82 -----GlyHis-----AsnArgMetTrpArg 88
Db 829 CTTCAGTCTTTAAACATAGATGTTATATCTACACTTGGACAACTGATGCTCTTTGGAGG 888
```

```
QY 89 LysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
889 AAATCCGTTACCCATAATAATGGCACATGTTTGGGACGGATCTCAATCGAAATTC 948
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerCysSerGluThrTrpCys 128
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 949 ---AATGCATCTGGTGTAGTATTGGTGCTCTAGAACTGCCAAGATCAAACTCTGT 1005
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1006 GGCACAGGGCCAGTGTCTGAACACAGAGACTAAAGCTGTTGCCAGCTTCATAGAGCAAG 1065
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1066 AAGGATGATATTTTGGTCTCCGACCATCATCTTATGGGCGATTAATTCACACCT 1125
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1126 TACGGCTACACCAAAATAAATCAAGTAACCCACCAAGAAATGATTCAAGTTGCACAGAAG 1185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1186 GCAGCAAAATGCATG---AAAGCAAAAGTATGGAACCAATATAGAGTTGGATCGAGTGCA 1242
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyrAspLeuGlyIleLys 228
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1243 GATATTTTATATCCCTCAGGGTCTTCAAGAGATTGGGCCCGAGACATTTGGGATTC 1302
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 229 TyrSerPheThr 232
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1303 TTCTCATACG 1314
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 12
US-10-176-306-73
; Sequence 73, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Bandaru, Rajasekhara
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
; FILE REFERENCE: 10448-195001
; CURRENT APPLICATION NUMBER: US/10/176,306
; CURRENT FILING DATE: 2002-06-20
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 10/001,137
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: PCT/US01/45291
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/248,362
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,331
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,365
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/250,077
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,327
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,176
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 10/023,617
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/49416
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,249
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,405
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 10/083,248
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/US01/46717
```

```
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242,324
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,518
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 1603
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (272)...(1603)
US-10-176-306-73

Alignment Scores:
Pred. No.: 3,45e-63 Length: 1603
Score: 550.50 Matches: 104
Percent Similarity: 56.44% Conservative: 45
Best Local Similarity: 39.39% Mismatches: 76
Query Match: 41.14% Indels: 39
DB: 16 Gaps: 4

US-09-980-881A-4 (1-246) x US-10-176-306-73 (1-1603)

QY 6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPileGluPheIleThrGluArg 25
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 800 TATAACATATACACCCCATGGAGAGATCTATGAGTGGATGAGAGATCAGTGAGAAG 859
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 860 TACAAGGAAGTGGTGACACAGCATTTCTTAGGAGTGACCTATGAGACCCACCCATATAT 919
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpLeuAspCysGly 65
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 920 TATCTGAAGATCAGCCCAACCATCTGTAATCCCAAGAAAATCATTTGGATGGACTGTGA 979
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTyrPheIle 81
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 980 ATTACGCCAGAGATGGATTGCTCTGCTTTTGGCAATGGTTCGTAAGAAATCTTA 1039
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 81 81
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1040 CAAAACATAAAGACAACTCAAGGATACCGAAGCTCTTAGGAACCTGGACTTCTATGTC 1099
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 82 82
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1100 CTTCCAGTTCTTAACATAGATGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGG 1159
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 89 LysAsnArgSerPheTyrAlaAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1160 AAATCCCGTTCAACCCTAATTAATGGACATGTTTGGGACGGATCTCAATCGAAATTC 1219
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 109 AlaSerLysHisTrpCysGluGluAlaSerSerSerCysSerGluThrTyrCys 128
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1220 ---AATGCATCTGGTGTAGTATTGGTGCTCTAGAACTGCCAAGATCAAACTCTGT 1276
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAsn 148
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1277 GGCACAGGGCCAGTGTCTGAACACAGAGACTAAAGCTGTTGCCAGCTTCATAGAGCAAG 1336
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1337 AAGGATGATATTTTGGTCTCTGACCATGCACTCTTATGGGCGATTAATTCACACCT 1396
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1397 TACGGCTACACCAAAATAAATCAAGTAACCCAGAAATGATTCAGAGTGACAGAG 1456
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 1457 GCAGCAATGCAATG---AAAGCAAGTATGGAACCAATATATAGAGTTGGATCGAGTGCA 1513
Qy 209 GluThrLeuTyrLeuAlaProGlyGlyAlaAspTrpIleTyrAspLeuGlyIleLys 228
Db 1514 GATATTTTATATGCTCATCAGGGTCTTCAAGAGATTGGCCCGGACGACATTTGGGATTCCC 1573
Qy 229 TyrSerPheThr 232
Db 1574 TTCTCATATACG 1585

RESULT 13
US-09-910-059-112
; Sequence 112, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; TITLE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; CURRENT FILING DATE: 2001-07-23
; PRIOR FILING DATE: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 112
; LENGTH: 1870
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: expected PCR insert
US-09-910-059-112

Alignment Scores:
Pred. No.: 7,07e-63 Length: 1870
Score: 549.00 Matches: 108
Percent Similarity: 57.62% Conservative: 47
Best Local Similarity: 40.15% Mismatches: 74
Query Match: 41.03% Indels: 40
DB: 10 Gaps: 5

US-09-980-881A-4 (1-246) x US-09-910-059-112 (1-1870)
Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 934 GCAACAGGACACAGTATGAGAGTACAAACAGTGGGAAACGATAGAGCGCTGACTCA 993
Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 994 CAAGTCGCCACTGAGATCCAGCGCTCATCTCGCAGTGTATCGGAACACCATTTGAG 1053
Qy 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 1054 GGAGCGCTATTATCTCTCTGAGGTT---GGCAAGCTGGACAAATAAGCTGCGATT 1110
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 1111 TTCATGCACTGTGGTTTCCATCCAGAGTGGATTCTCTGCTATCTGCGAGTGGTTT 1170
Qy 81 Ile----- 81
Db 1171 GTAAGAGAGGCTGTTCTGATCTATGAGCTGAGATCCAAAGTGACAGAGCTTCTCGACAAG 1230
Qy 82 -----GlyHis----- 83
```

```
Db 1231 TTAGACTTTTATGCTCTGCTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 1290
Qy 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 1291 AGCCGATTTTGGAGAAAGACTCGCTCCACCTACTGGATCTAGCTGCAATTGGCACAGAC 1350
Qy 104 LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGluAlaSerSerSerCys 123
Db 1351 CCCAACAGAAATTT---GATGCTGGTTGGTGTGAATTTGGAGCCCTCTCGAAACCCCTGT 1407
Qy 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 1408 GATGAAACTTACTGTGGACCTGCCGAGAGTCTGAAAGAGAGACCAAGCCCTGGCTGAT 1467
Qy 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyrIleSerMethHisSerTyrSerGln 163
Db 1468 TTATCCGCAACAAACTCTCTCCATCAAGGATATCTGACAAATCCACTCGTACTCCCAA 1527
Qy 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 1528 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAGAACAAATGCTGAGTTGAAT 1587
Qy 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1588 GCCTGGCTAAAGCTACTGTGAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1644
Qy 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
Db 1645 TATGGCCCGGAGCTACAAACAATCTATCTCTGCTGGGACTTCTAAAGAGCTGGGCTTAT 1704
Qy 224 AspLeuGlyIleLysTyrSerPheThr 232
Db 1705 GACCAAGAAATCAGATATTCTTCACC 1731

RESULT 14
US-09-910-059-124
; Sequence 124, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; TITLE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; CURRENT FILING DATE: 2001-07-23
; PRIOR FILING DATE: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 124
; LENGTH: 2154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: humanised pre-pro HCFB-linker-Fd sequence
US-09-910-059-124

Alignment Scores:
Pred. No.: 8,91e-63 Length: 2154
Score: 549.00 Matches: 108
Percent Similarity: 57.62% Conservative: 47
Best Local Similarity: 40.15% Mismatches: 74
Query Match: 41.03% Indels: 40
DB: 10 Gaps: 5
```

```
US-09-980-881A-4 (1-246) x US-09-910-059-124 (1-2154)
QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 331 GCAACAGGACACAGATTATGAGAGTACAAAGTGGAAACGATAGAGGCTTGGACTCAA 390
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 391 CAAGTCGCCACTGAGATCCAGCCCTCATCTCTCGCAGTGTATATCGGAACCACTTTGAG 450
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 451 GGACGGCTATTATCTCTGAGGTT---GGCAAGCTGGACAAATAAGCCTGCCATT 507
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 508 TTCATGAGCTGTGTTTCCATGCCAGAGTGGATTCTCTCGCATCTCTCGCAGTGGTT 567
QY 81 Ile----- 81
Db 568 GTAAGAGAGGCTGTTCTGTACTATGACGCTGAGATCCAAAGTGACAGAGCTTCTCGACAAG 627
QY 82 -----GlyHis----- 83
Db 628 TTAGACTTTATGCTGCTGCTGTGCTCAATATTGATGGCTTACATCTACCTGGACCAAG 687
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 688 AGCCGATTTGGAGAAAGACTCGCTCCACCATACTGGATCTAGCTGCAATGGCAGAC 747
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 748 CCCAACAGAAATTTT---GATGCTGTTGTGTGTAATTTGGAGCCTCTCGAAACCCCTGT 804
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 805 GATGAACACTTACTGTGGACCTGCCGAGAGTCTGNAAGAGAGACCAGGCCCTGGCTGAT 864
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 865 TTCATCGGCACAAACTCTCTCCATCAGGCATATCTGACATCCACTGCTACTCCAA 924
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 925 ATGATGATCTACCTTACTCATATGCTTACAAACTCGGTGAGAACAACTGTGAGTTGAAT 984
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 985 GCCTCGCTAAAGCTACTGTGAAGAAGACTT---GCCTCACTGCACGCGCACCAAGTACACA 1041
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
Db 1042 TATGCGCCGGAGGACTCAACAATCTATCTCTGCTGGGACTTCTAAAGACTGGGCTTAT 1101
QY 224 AspLeuGlyIleLysTyrSerPheThr 232
Db 1102 GACCAAGGAATCAGATATCTTCCACC 1128

RESULT 15
US-09-954-456-1141
; Sequence 1141, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
```

```
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1141
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-954-456-1141

Alignment Scores:
Pred. No.: 6,45e-63 Length: 1332
Score: 547.50 Matches: 110
Percent Similarity: 57.99% Conservative: 46
Best Local Similarity: 40.89% Mismatches: 72
Query Match: 40.92% Indels: 41
DB: 10 Gaps: 6

US-09-980-881A-4 (1-246) x US-09-954-456-1141 (1-1332)
QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 354 GCAACAGGACACAGATTATGAGAGTACAAAGTGGAAACGATAGAGGCTTGGACTCAA 413
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 414 CAAGTCGCCACTGAGATCCAGCCCTCATCTCTCGCAGTGTATCGGAACCACTTTGAG 473
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 474 GGACGGCTATTATCTCTGAGGTT---GGCAAGCTGGACAAATAAGCCTGCCATT 530
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 531 TTCATGAGCTGTGTTTCCATGCCAGAGTGGATTCTCTCGCATTTCTGCCAGTGGTTT 590
QY 81 Ile----- 81
Db 591 GTAAGAGAGGCTGTTCTGTACTATGACGCTGAGATCCAAAGTGACAGAGCTTCTCGACAAG 650
QY 82 -----GlyHis----- 83
Db 651 TTAGACTTTATGCTGCTGCTGCTCAATATTGATGGCTTACATCTACCTGGACCAAG 710
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 711 AGCCGATTTGGAGAAAGACTCGCTCCACCATACTGGATCT---AGCATTTGGCAGAC 767
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 768 CCCAACAGAAATTTT---GATGCTGTTGTGTGTAATTTGGAGCCTCTCGAAACCCCTGT 824
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 825 GATGAACACTTACTGTGGACCTGCCGAGAGTCTGNAAGAGAGACCAGGCCCTGGCTGAT 884
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
```

Db 885 TTATCCGCCAACAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCA 944
QY 164 HistileValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 945 ATGATGATCTACCCCTTACTCATATGCTTACAACACTCGTGAGAACAAATGCTGAGTTGAAT 1004
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1005 GCCCTGGCTAAAGCTACTGTGAAGAAGACTT--GCCTCACTGCACGGCACCAGTACACA 1061
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
Db 1062 TATGCCCGGGAGCTACACAATCTATCTGCTGCTGGGGCTCTGACGACTGGGCTTAT 1121
QY 224 AspLeuGlyIleLysTyrSerPheThr 232
Db 1122 GACCAGGAATCAGATATTCCTTCACC 1148

Search completed: January 2, 2004, 10:05:26
Job time : 613.309 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:47:17 ; Search time 2179.08 Seconds
(without alignments)
2743.777 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQHSLNEIYSWIE.....IKYFTSNPPVEXLLPLSLK 246

Scoring table:
BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09980881/runat_31122003_135916_26873/app_query.fasta_1.1628
-DB=EST -QPM=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881 @CGN 1.1 5167 @runat_31122003_135916_26873 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_man:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_pmg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1112	83.1	1446	11	AK004045	AK004045 Mus muscu
2	978	73.1	852	10	BG569281	BG569281 602588573
3	831.5	62.1	688	9	AV651709	AV651709 AV651709
4	828.5	61.9	806	10	BG567264	BG567264 602589745
5	827.5	61.8	577	14	CB154355	CB154355 K-EST0212
6	812.5	60.7	973	9	AL536068	AL536068 AL536068
7	795	59.4	510	14	CB156851	CB156851 K-EST0215
8	793.5	59.3	971	9	AL536069	AL536069 AL536069
9	732	54.7	434	14	CB164014	CB164014 K-EST0225
10	698.5	52.2	723	9	AW413090	AW413090 UQ52C03.X
11	685.5	51.2	772	9	AI255929	AI255929 ui83d11.x
12	676.5	50.6	692	10	BF384322	BF384322 602046670
13	659.5	49.3	747	9	AI255397	AI255397 ui94a04.x
14	656.5	49.1	713	12	BI246210	BI246210 602958724
15	652.5	48.8	505	14	CB161965	CB161965 K-EST0222
16	649	48.5	645	10	BG642020	BG642020 pgl1c.pk0
17	631	47.2	546	14	H61865	H61865 yr02b04.r1
18	630	47.1	756	9	AI595707	AI595707 uj68d08.x
19	624	46.6	747	10	BG565100	BG565100 602583720
20	616.5	46.1	932	13	EX451532	EX451532 BX451532
21	607	45.4	378	9	AV651204	AV651204 AV651204
22	607	45.4	1041	11	AK008678	AK008678 Mus muscu
23	606	45.3	853	12	BI144644	BI144644 602909932
24	589	44.0	846	13	BU488930	BU488930 603972157
25	581.5	43.5	927	13	BU487280	BU487280 603972169
26	580.5	43.4	1816	11	AK078883	AK078883 Mus muscu
27	578	43.2	750	13	BU487283	BU487283 604126713
28	577.5	43.2	1429	11	AK003061	AK003061 Mus muscu
29	565	42.2	736	13	BU488223	BU488223 604127073
30	563.5	42.1	1004	14	BY704293	BY704293 BY704293
31	563	42.1	714	13	BU488385	BU488385 604124863
32	561.5	42.0	843	10	BG216526	BG216526 RST36106
33	561	41.9	1176	9	AB030423	AB030423 AB030423
34	561	41.9	1453	11	AK088906	AK088906 Mus muscu
35	560.5	41.9	466	9	AA244760	AA244760 mx05d08.r
36	560	41.9	897	14	CB202958	CB202958 AGENCOURT
37	555.5	41.5	984	14	CB207599	CB207599 AGENCOURT
38	553	41.3	766	13	BU490586	BU490586 604131409
39	552	41.3	820	14	CA776082	CA776082 1099d12.y
40	550	41.1	871	13	BU416594	BU416594 603670981
41	549	41.0	864	14	CB755833	CB755833 AGENCOURT
42	548	41.0	809	13	BU488219	BU488219 604126116
43	547.5	40.9	805	13	BU487175	BU487175 604126902
44	547	40.9	616	12	BI064948	BI064948 pgl1n.pk0
45	547	40.9	797	13	BU489339	BU489339 604131182

ALIGNMENTS

RESULT 1
AK004045
LOCUS
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length
enriched library, clone:110032P04 product:carboxypeptidase B2
(plasma), full insert sequence.
ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus
ORGANISM Mus musculus (house mouse)
Mammalia; Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Mollusca; Sciuromorphi; Muridae; Murinae; Mus.

QY 61 TrolleaspCysGlyIleHisAlaArgGluTrrPheSerProAlaPheCysLeuTrrPhe 80
 Db 561 TGGATCGACTGTGGAATCCATGCCAGAGATGGATTTCACCTGCTTCTGTTTGGTTC 620
 QY 81 IleglyHis----- 83
 Db 621 ATAGGCTACGTGACACAATTCATGGGAAGAAATCTGTATACAGACTTCTGAGGCAC 680
 QY 83 ----- 83
 Db 681 GTGGATTTTACATCATGCCGCTGATCAACGTCGATGGCTATGACTACACGTGGAAAAAG 740
 QY 84 AsnArgMetTrrArgLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAsp 103
 Db 741 AATCGAATGTGGAGGAAGAACCTCTCTGCTCAAGAACCAACCGCTGCGGGGACAGAC 800
 QY 104 LeuAsnArgAsnPheAlaSerLysHisTrrCysGluGluGlyAlaSerSerSerCys 123
 Db 801 CTGAACAGGAACCTTCGCTTCCAAACACTGGTGAGAAAGGTGCTCAAGTTCCTCTGC 860
 QY 124 SerGluThrTrrCysGlyLeuTyProGluSerGluProGluValLysAlaValAlaSer 143
 Db 861 TCTGAACCTTACTGTGACTTTATCTCTGAGTCTGAGCCAGGCTGAAGCAGTGGCTGAC 920
 QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyIleSerMetHisSerTySerGln 163
 Db 921 TCTTGAAGAAATATCGACCACTTAAGCTTATCATCATGATGACTCATCATCTCCAA 980
 QY 164 HisIleValPheProTySerTyThrArgSerLysSerLysAspHisGluGluLeuSer 183
 Db 981 CAATATCTGTTCCCTATTCCTATACAGAGCAAGCAAGCAAGCAAGCAAGCAAGTCT 1040
 QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyThr 203
 Db 1041 CTAGTGGCCAGCAAGCAGTTCGTGCAATTTGAAAGTATTATATAAACACCAAGGTACACA 1100
 QY 204 HisGlyHisGlySerGluThrLeuTyLeuAlaProGlyGlyCysAspTrrPheTy 223
 Db 1101 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATGGATCTAT 1160
 QY 224 AspLeuGlyIleLysTySerPhe----- 231
 Db 1161 GATTGGGCATCAATATTCGTTTACATTTGAGCTCGAGATACAGCAGATACGATTC 1220
 QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuTy 246
 Db 1221 TTGCTGCTGAGAGATACATCAACCCACTTGTGTCAGAGCTTTGCCGCCCTCTTAA 1280
 QY 246 s 246
 Db 1281 A 1281

RESULT 2
 BG569281
 LOCUS 852 bp mRNA linear EST 10-APR-2001
 DEFINITION 602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5', mRNA sequence.
 ACCESSION BG569281
 VERSION BG569281.1 GI:13576934
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 852)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLCM1580 row: b column: 19
 High quality sequence stop: 751.
 FEATURES
 source
 1..852
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4722354"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_76"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site1: SfiI (ggccattatggc); Site2: SfiI (ggccattatggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGAGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."
 BASE COUNT 251 a 177 c 187 g 237 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6,15e-101 Length: 852
 Score: 978.00 Matches: 197
 Percent Similarity: 76.15% Conservative: 1
 Best Local Similarity: 75.77% Mismatches: 5
 Query Match: 73.09% Indels: 58
 DB: 10 Gaps: 2

US-09-980-881A-4 (1-246) x BG569281 (1-852)

QY 43 ProLeuTyValLeuLysValSerGlyLysGluInThrAlaLysAsnAlaIleTrrPhe 62
 Db 2 CCACCTCTATGTTTAAAGGTTTCTGGAAAGAAACAGCAGCCCAAAATGCCATATGGATT 61
 QY 63 AspCysGlyIleHisAlaArgGluTrrPheSerProAlaPheCysLeuTrrPheIleGly 82
 Db 62 GACTGTGGAAATCCATGCCAGAGATGGATCTCTCTGCTTCTGCTTGTGGTTTCATAGGC 131
 QY 83 His----- 83
 Db 122 CATATAACTCAATCTTATGGGATAATAGGCAATATACCAATCTCTGAGGCTTGTGGAT 181
 QY 84 -----AsnArg 85
 Db 182 TTCTATGTTTATGCCAGTGGTAAATGTGGATGGTTATGACTACTCATCGAAAGAAATCGA 241
 QY 86 MetTrrArgLysAsnArgSerPheTyAlaAsnAsnHisCysIleGlyThrAspLeuAs 105
 Db 242 ATGTGGACAAAGAACCGTTCTTCTATCGCAACATCATTCATCGAACACAGACCTGAA 301
 QY 105 nArgAsnPheAlaSerLysHisTrrCysGluGluGlyAlaSerSerSerCysSerGly 125
 Db 302 TAGGAACCTTGTCTCCAAACACTGGTGTGAGGAGGTGATCCAGTTCCTCATGCTCGGA 361
 QY 125 uThrTyCysGlyLeuTyProGluSerGluProGluValLysAlaValAlaSerPheLe 145
 Db 362 AACCTACTGTGACTTTTATCTCTGAGTCAGAACCAAGGAGGAGGAGTGGCTAGTTTCTT 421
 QY 145 uArgArgAsnIleAsnGlnIleLysAlaTyIleSerMetHisSerTySerGlnHisI 165
 Db 422 GAGAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCATCTCCAGCATAT 481
 QY 165 eValPheProTySerTyThrArgSerLysSerLysAspHisGluGluLeuSerLeuVa 185
 Db 482 AGTGTTTCCATATTCCTATACAGAAAGTAAAGCAAGACCATGAGGAACTGTCTCTAGT 541


```

QY 185 lalaSerGluAlaValAlaGlaLeGluYsThrSerLysAsnThrArgTyrThrHisG1 205
Db 542 AGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTTAGTAAATAACACAGGTATACATGC 601
QY 205 yHisGlySerGluThrLeuTyrIleuAlaProGlyGlyAspAspTrpIleTyr-AspL 225
Db 602 GCATGGCTCGAAGACCTTATACCTAGCTCTCGAGGTGGGAGCAGATTGGATCTATGCATT 661
QY 225 euGlyIleLysTyrSerPhe----- 231
Db 662 TGGGCATCAGATATTCGTTTACAAATGGAACCTTCGAGATACGGGCACATACCGATTCTT 721
QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeu 245
Db 722 GCTGCCGAGCGTTACATCAAA-CCACCTGTAGAGAGCTTTTGGCGCTCTCTA 776

RESULT 3
AV651709 688 bp mRNA linear EST 15-JAN-2002
LOCUS AV651709 GLC Homo sapiens cDNA clone GLCCSF10 3', mRNA sequence.
DEFINITION AV651709
ACCESSION AV651709
VERSION AV651709.1 GI:9872723
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 688)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1. 688
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCCSF10"
/tissue_types="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 208 a 145 c 151 g 184 t
ORIGIN
Alignment Scores:
Pred. No.: 2,4e-84 Length: 688
Score: 831.50 Matches: 162
Percent Similarity: 89.50% Conservative: 0
Best Local Similarity: 89.50% Mismatches: 1
Query Match: 62.14% Indels: 18
DB: Gaps: 1

US-09-980-881a-4 (1-246) x AV651709 (1-688)

QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103

```

```

Db 46 AATCGAATGGGAGAAAGAACCGTCTTCTTATCGCAACAATCATTCGTCGGAACAGAC 105
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 106 CTGAATAGGAACCTTTCTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 165
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 166 TCGGAACCTACTGTGCACTTTATCTCTAGTCAGAACCAAGAGTGAAGCAGTGGCTAGT 225
QY 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyrIleSerMethHisSerTyrSerGln 163
Db 226 TTCTTGAGAAGAAATATCAACCAAGATTAAAGCATATACATCAGCATGCATTCATCTCCAG 285
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 193
Db 286 CATATAGTGTTCATATTCCTATATACCAAGTAAAGCAAGCAACCATGAGGAACCTGTCT 345
QY 184 LeuValAlaSerGluAlaValAlaArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 346 CTAGTAGCAGCAGTGAACAGCTTCGTGCTATTGAGAAAACCTAGTAAATAACCAAGTATACA 405
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyr 223
Db 406 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGGTGGGACGATTGATCTAT 465
QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 466 GATTTGGGCATCAAAATATTTCGTTTACAATGGAACCTCGAGATACCGGCACATACCGATTTC 525
QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeu 246
Db 526 TTTGTGGCGGAGCGTTACATCAAAACCACTGTTGAGAAAGCTTTTGGCGCTGTCTTTAA 585
QY 246 s 246
Db 586 A 586

RESULT 4
BG567264 806 bp mRNA linear EST 10-APR-2001
LOCUS BG567264 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723690 5',
DEFINITION mRNA sequence.
ACCESSION BG567264
VERSION BG567264.1 GI:13574917
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 806)
AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1583 row: j column: 11
High quality sequence stop: 772.
Location/Qualifiers
1. 806
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4723690"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"

```

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgccg); Site 2: SfiI (ggccattagcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGCGCATTTAGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCCGACATC-dT(30)BN-3' where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."

BASE COUNT	239 a	158 c	162 g	247 t
ORIGIN				

Alignment Scores:	
Pred. No.:	6,78e-84
Score:	828.50
Percent Similarity:	89.94%
Best Local Similarity:	89.94%
Query Match:	11.92%
DB:	10
Length:	806
Matches:	161
Conservative:	0
Mismatches:	0
Indels:	18
Gaps:	1

US-09-980-881A-4 (1-246) x BG567264 (1-806)

Qy	86	MetTrrArgLysAsnArgSerPheTyrrAlaIasnHHisCysIleGlyThrAspLeuAsn	105
Db	2	ATGTGAGAGAAACCGTCTCTTCTATCGGAACAATCATTTGCATCGGAAACAGACCTGAAT	61
Qy	106	ArgAsnPheAlaSerLysHisTrpCysGluGluAlaSerSerSerCysSerGlu	125
Db	62	AGGAACATTGCTTCCAAACACTGGTGTCAGAAAGTGCAATTCCTCATGCTCGGAA	121

Oy	126	ThrTyrCysGlyLeuTyrProGluSerGluProGluValIysAlaValAlaSerPheLeu	145
Db	122	ACCTACTGTGGAGCTTTATCTCGTCAGAACCAAGTGGAAGGCAGTGGCTAGTTCCTTG	181
Oy	146	ArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMethHisSerTyrSerGlnHisIle	165
Db	182	ACAGAGAAATATATATATCCAGATTATACCATATATCTTCAGCTATGATATCTTCCTCCGCTATTA	241

[illegible]

Qy	206	HisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeu	225
Db	362	CATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGACGATTGGATCATGATTG	421
Qy	226	GlyIleLeuTyrSerPhe	231

Qy	232	-----	ThrSerAenProValGluLysIeuLeuProtenSerIeuLys	246
Db	482	CGGAGCGTTACATCAAA	CCACCCTGTAGAGACTTTCGCCGTCTCTTAANA	536
 RESULT 5 CB154355				

DEFINITION	K-EST0212246 L17N670205 Homo sapiens cDNA clone L17N670205-10-E10
ACCESSION	S1, mRNA sequence.
VERSION	CB154355
KEYWORDS	CB154355.1 GI:28139354
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Homo sapiens

REFERENCE

1 (bases 1 to 577)

Aukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Kim.N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R. AUTHORS

Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 10 row: E column: 10
High quality sequence stop: 577

FEATURES
SOURCE

```
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="L17N670205-10-E10"  
/sex="F"  
/lab_host="Top10F"  
/clone_lib="L17N670205"
```

Site-2; NotI. The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."

BASE COUNT	175 a	122 c	128 g	152 t
ORIGIN				

Alignment Scores:					
Pred. No.:	5, 23e-84	Length:	577		
Score:	827.50	Matches:	161		
Percent Similarity:	89.44%	Conservative:	0		
Best Local Similarity:	89.44%	Mismatches:	1		
Query Match:	61.85%	Indels:	18		
DB:	14	Gaps:	1		

US-09-980-881A-4 (1-246) x CB154355 (1-577)

Qy	85	ArgMetTrpArgIysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspIeu	104
Db	1	CGAATGGGAGAAAGAACCGTCTTTCTATCGCAACAATCATTTGCATCGGAACACACCTTG	60
Qy	105	AsnArgAsnPheAlaSerLysHisTyrCysGluGluGlyAlaSerSerSerSerCysSer	124
Db	61	AATAGGAACATTTGGTTCCAAACATCGTGTGAGGAAGGTGCATCAGTTCTCTCATGCTCG	120

Qy	125	GlutThrTyrCysGlyLeuTyrProGluUserGluProGluValLeuAlaValalaSerPhe	144
Db	121	GAACCTTACTGTGTGACCTTTATCTCTGAGTCAGAACGAGGCGATGGCTAGTTTC	180
Qy	145	LeuArgArgAsnIleAsnGlnIleLeuAlaTyrIleSerMetHisSerTyr-SerGlnHis	164
Db	181	TTGAGAGAGAAATATCAACGAGATTAAAGCATATCATCAGCATGCTTATATATCCACGAT	240

Oy 165 IleValPheProTyrSerTyrThrArgSerLysSeryAspHisGluGlucLeuSerTeu 184
Dd 241 ATAGTGTTCCATATTCTGTACACGAAGTAAGAACAACCATGAGCAACTGTCCTA 300
Oy 185 ValAlaSerGluAlaValArgAlaIleGlyLysThrSerLysAsnThrArgTyrrThrHis 204
Dd 291 CTATCGCNCNCTGGGCCGCCTGCTCTTGTGAGCAAATAATCATAAATTCAGCTCTTCACACT 360

Qy	205	GlyHisGlySerGluThrLeuTy ^r LeuAlaProGlyGlyGlyAspAspT ^r ileTyrAsp	224
Db <td>361 <td>GGCATGGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGACGATTGGATCTATCAT</td> <td>420</td> </td>	361 <td>GGCATGGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGACGATTGGATCTATCAT</td> <td>420</td>	GGCATGGCTCAGAAACCTTATACCTAGCTCTGGAGTGGGACGATTGGATCTATCAT	420
Qy <td>225 <td>LeuGlyIleIysTyrSerPhe-----</td> <td>231</td> </td>	225 <td>LeuGlyIleIysTyrSerPhe-----</td> <td>231</td>	LeuGlyIleIysTyrSerPhe-----	231

```

QY 232 -----ThrsAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
Db 481 CTGCGGAGCGTTACATCAACCCACCTGTAGAGAGCTTTTGCCTGCTCTATAAA 538

RESULT 6
AL536068/c
LOCUS
DEFINITION
AL536068 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF022YE21 3-PRIME, mRNA sequence.
ACCESSION
AL536068
VERSION
AL536068.2 GI:31260938
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 973)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12799561.
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5944.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF022AC11NP1&cluster=6944.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0DF022AC11NP1.
LOCATION/Qualifiers
1. 973
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 284 a 193 c 188 g 250 t 58 others
ORIGIN

Alignment Scores:
Pred. No.: 6.13e-82 Length: 973
Score: 812.50 Matches: 159
Percent Similarity: 88.46% Conservative: 2
Best Local Similarity: 87.36% Mismatches: 2
Query Match: 60.72% Indels: 19
DB: 9 Gaps: 1

US-09-980-881A-4 (1-246) x AL536068 (1-973)

QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 954 ATCTGATGTGGAGAAAGACCGTCTCTTATGCGGACATCATGTGATGGACAGAC 895

104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 894 CTGAATAGGAACCTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 835

124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 834 TCGGAAACCTACTGTGACATTTATCCCTGAGTCAGACCAAGAGTGAGCGGCTAGT 775

144 PheLeuArgAsnIle-AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerG1 163

```

```

Db 774 TTCTTGAGAGAAATATCAAGCAGCATTAAGCATACATCAGCATGCAATTCATACCTCCA 715
QY 163 nhisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSe 183
Db 714 GCATATAGTGTTCCTCATATTCCTATACACAGAAAGTAAAGCAAGACCATGAGGAACGTGC 655
QY 183 rleuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrTh 203
Db 654 TCTAGTAGCCAGTGAAGCAGTTCGTCTATTGAGAAAACTAGTAAATAATACAGGTATAC 595
QY 203 rHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTy 223
Db 594 ACATGCCCATGCTCAGAAACCTTATACCTAGTCTCTGGAGGTGGGGACGATGGATCTTA 535
QY 223 rAspleuGlyIleLysTyrSerPheThr----- 232
Db 534 TGATTTGGGCATCAATTTTTCGTTTACAAATTGAACTTCGAGTTACGGGCACATACGGATT 475
QY 233 -----SerAsnProProValGluLysLeuLeuProLeuSerLeuL 246
Db 474 CTGTGTCGGGGCGGCTTCTCAACCCCTGTAGAGAGCTTTTGGCGCTGTCTCTAA 415
QY 246 YS 246
Db 414 AA 413

RESULT 7
CB156851
LOCUS
DEFINITION
K-EST0215830 L17N670205n1 Homo sapiens cDNA clone
L17N670205n1-4-H06 5', mRNA sequence.
ACCESSION
CB156851
VERSION
CB156851.1 GI:28141979
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 510)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 4 row: H column: 06
High quality sequence stop: 510.
FEATURES
source
1. 510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-4-H06"
/sex="F"
/lab_host="Top10F"
/clone_lib="L17N670205n1"
/notes="Organ: liver; Vector: pTTT3-Pac; Site_1: EcoRI;
Site_2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."
BASE COUNT 158 a 109 c 113 g 130 t
ORIGIN

Alignment Scores:

```

```

Pred. No.: 2,23e-80 Length: 510
Score: 795.00 Matches: 147
Percent Similarity: 99.32% Conservative: 0
Best Local Similarity: 99.32% Mismatches: 1
Query Match: 59.42% Indels: 0
DB: 14 Gaps: 0

US-09-980-881A-4 (1-246) x CBI56851 (1-510)

QY 85 ArgMetTrrArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 104
Db 1 CGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACCAATCAATGTCATCGGACAGACCTG 60

QY 105 AsnArgAsnPheAlaSerLysHisTrrCysGluGluGlyAlaSerSerSerCysSer 124
Db 61 AATAGGAACCTTTGCTTCCAAACACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 120

QY 125 GluThrTrrCysGlyLeuTrrProGluSerGluProGluValLysAlaValAlaSerPhe 144
Db 121 GAAACCTACTGTGGACTTTATCTCTAGTCTAGAACCAAGAGTGAAGCACTGGCTAGTTTC 180

QY 145 LeuArgArgAsnIleAsnGlnIleLysAlaTrrLysSerMetHisSerTrrSerGlnHis 164
Db 181 TTGAGAAGAAATATCAACAGATTAAGCATATACATCAGCATGCAATTCATCTACCTCCAGCAT 240

QY 165 IleValPheProTrrSerTrrThrArgSerLysSerLysAspHisGluLysLeuSerLeu 184
Db 241 ATAGTGTTTCCATATTCCTATACAGAGTAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 300

QY 185 ValAlaSerGluAlaValAlaGlnIleLysAlaTrrLysSerLysAsnThrArgTrrThrHis 204
Db 301 GTAGCCAGTGAAGCAGTTCGTCTATTGAGAAATAGTAAAGCAAGCAAGCAAGCAAGCAAG 360

QY 205 GlyHisGlySerGluThrLeuTrrLeuAlaProGlyGlyGlyAspTrrPheTrrAsp 224
Db 361 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGACGATTGGATCATGAT 420

QY 225 LeuGlyIleLysTrrSerPheThr 232
Db 421 TTGGGCATCAAAATATTCGTTTACA 444

RESULT 8
AL536069 971 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION
AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF022YE21 5-PRIME, mRNA sequence.
AL536069
AL536069.2 GI:31260939
EST.
Homo sapiens (human)
ORGANISM
Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 971)
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12799562..
Contact: Genoscope
Genoscope - Centre National de Sequenage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6944.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF022AC11Q1&cluster=6944.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF022AC11Q1.
Location/Qualifiers
i. .971
/organism="Homo sapiens".

FEATURES
source
CBI64014 434 bp mRNA linear EST 30-JAN-2003
K-EST0225031 L17N670205n1 Homo sapiens cDNA clone
L17N670205n1-38-G03 5', mRNA sequence.
CBI64014
ACCESSION
VERSION CBI64014.1 GI:28150140
KEYWORDS
SOURCE Homo sapiens (human)

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

BASE COUNT 282 a 201 c 211 g 272 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 9,05e-80 Length: 971
Score: 793.50 Matches: 148
Percent Similarity: 79.14% Conservative: 0
Best Local Similarity: 79.14% Mismatches: 2
Query Match: 59.30% Indels: 37
DB: 9 Gaps: 1

US-09-980-881A-4 (1-246) x AL536069 (1-971)

QY 1 AlaSerAlaSerTrrTrrGluGlnTrrHisSerLeuAsnGluLysTrrSerTrrPheGlu 20
Db 407 GCCTCCGATCGTACTATGTAACAGATATCACTCACTAAATGAATCTATTTCTTGGATAGAA 466

QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 467 TTTATTAACGTAGAGAGCATCTCTGATATGTTTCAAAATCCACATTTGGATCTCTCATTTGAG 526

QY 41 LysTrrProLeuTrrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 527 AAGTACCCACTCTATGTTTAAAGGTTTCTGAAAGAACAGAGCCAAATGCCATA 586

QY 61 TrrPheAspCysGlyIleHisAlaArgGluTrrPheSerProAlaPheCysLeuTrrPhe 80
Db 587 TGGATTGACTGTGGAATCCATGCCAGAGATGATCTCTCTGCTTCTCTGCTTGTGGTTTC 646

QY 81 IleGlyHis----- 83
Db 647 ATAGGCCATATTAACCTCAATTTCTATGGGATAATAGGCAATATACCAATCTCTGAGGCTT 706

QY 83 ----- 83
Db 707 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAG 766

QY 84 AsnArgMetTrrArgLysAsnArgSerPheTrrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 767 AATCGAATGTGGAGAAAGAACCGTTCTTCTATGCGAACCAATCAATGCAATGGACAGAC 826

QY 104 LeuAsnArgAsnPheAlaSerLysHisTrrCysGluGluGlyAlaSerSerSerCys 123
Db 827 CTGAATAGGAACCTTTGCTTCCAAACACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 886

QY 124 SerGluThrTrrCysGlyLeuTrrProGluSerGluProGluValLysAlaValAlaSer 143
Db 887 TCGGAACCTACTGTGGACTTTATCTGAGTCAGAACCAAGAGTGAAGCAAGTGGCTAGT 946

QY 144 PheLeuArgArgAsnIleAsn 150
Db 947 TTCTTTGAGAAGAAATTAAC 967

RESULT 9
CBI64014
LOCUS
DEFINITION
CBI64014
ACCESSION
VERSION CBI64014.1 GI:28150140
KEYWORDS
SOURCE Homo sapiens (human)

```



```

Db      296  GATTGGGCATCAATATTGGTTTACAAATTGAGCTCCGAGATACAGGCAGATACGGGATTC 237
Qy      232  -----ThrsSerAnProValGluLysLeuLeuProLeuSerLeuLys 246
Db      236  TTGCTGCTGAGAGATACATCAAAACCCACTTGTGCAGAGCTTTGGCGCCGATCTCTAA 177
Qy      246  s 246
Db      176  A 176

RESULT 12
BF384322      692 bp mRNA linear EST 27-NOV-2000
LOCUS      602046670F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4196129 5',
DEFINITION mRNA sequence.
ACCESSION BF384322
VERSION BF384322.1 GI:11365627
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9532 row: 1 column: 18
High quality sequence stop: 689.
FEATURES
source Location/Qualifiers
1..692
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4196129"
/lab_host="DRI0B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 211 a 156 c 139 g 186 t
ORIGIN

Alignment Scores:
Pred. No.: 1.24e-66 Length: 692
Score: 676.50 Matches: 134
Percent Similarity: 81.61% Conservative: 8
Best Local Similarity: 77.01% Mismatches: 14
Query Match: 50.56% Indels: 18
DB: 10 Gaps: 1

US-09-980-881A-4 (1-246) x BF384322 (1-692)

Qy      91  ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 110
Db      1  CGCTCTGCTCACAAGAACACACCGCTGGTGGGCACACGACCTGAACAGAACTTCGCTTCC 60
Qy      111  LysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeu 130
Db      61  AAACACTGGGTGAGAAAGTGGGTCAAGTTCCTCGCTCTGAAACCTACTGTGGACTT 120
Qy      131  TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 150

```

```

Db      121  TATCCTGAGTCTGAGCGACGAGGTGAAGCGAGTGGCTGACTTCTTGAGAGAAATATCGAC 180
Qy      151  GlnIleIysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 170
Db      181  CACATTAAAGCTTACATCATGATGCACTCATCTCCCAACAATATCTGTTTCCCTATTC 240
Qy      171  TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 190
Db      241  TATAACAGAGCAAGCAAGGACGACGAGAACTGTCTCTAGTGCCGACGAGCAGTT 300
Qy      191  ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 210
Db      301  CGTGCAATTGAAAGTATTAAATAAAACACACGAGTACACACGCGAGTGGCTCAGAAAGT 360
Qy      211  LeuTyrIleAlaProGlyGlyAspAspTyrIleTyrAspLeuGlyIleLysTyrSer 230
Db      361  TTATATCTAGCTCTCTGAGAGTTCGACGATTGGATCTATGATTGGGCATCAATATTCG 420
Qy      231  Phe-----ThrSe 233
Db      421  TTTTCAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCTGAGAGATACATC 480
Qy      233  rAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
Db      481  AAACCCACTGTGTCAGAGCTTTGGCGCCATCTCTATAAA 520

RESULT 13
AI255397      747 bp mRNA linear EST 12-NOV-1998
LOCUS      ui94a04.x1 Sugano mouse liver mlia Mus musculus cDNA clone
DEFINITION IMAGE:1890030 3' similar to TR:Q15114 Q15114 PREPRO-PLASMA
CARBOXYPEPTIDASE B. ; mRNA sequence.
ACCESSION AI255397
VERSION AI255397.1 GI:3862922
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 747)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:974354
Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 471.
FEATURES
source Location/Qualifiers
1..747
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1890030"
/sex="female"
/dev_stage="adult"
/lab_host="PH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site 1: DraIII
(CACTGTGTG); Site 2: DraIII (CACCATGTG); 1st strand cDNA

```



```

|||||
422 GTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCTCGAGATACAT 481
|||||
Qy 233 erAnProValGluLysLeuLeuProLeuSerLeuLys 246
|||||
Db 482 CAACCCACTTGTGAGAGAGCTTTGGCCGCACTCTTAAAA 522
|||||
RESULT 15
CB161965
LOCUS
DEFINITION K-EST0222418 L17N670205n1 Homo sapiens cDNA clone
L17N670205n1-15-E06 5', mRNA sequence.
CB161965
ACCESSION CB161965.1 GI:28148091
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 505)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 15 row: E column: 06
High quality sequence stop: 505.
Location/Qualifiers
FEATURES
source
1. 505
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-15-E06"
/sex="F"
/lab_host="Top10F'"
/clone_lib="L17N670205n1"
/notes="Organ: Liver; Vector: pT7T3-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."
BASE COUNT 160 a 99 c 108 g 138 t
ORIGIN
Alignment Scores:
Pred. No.: 4.18e-64 Length: 505
Score: 652.50 Matches: 125
Percent Similarity: 75.45% Conservative: 1
Best Local Similarity: 74.85% Mismatches: 4
Query Match: 48.77% Indels: 37
DB: 14 Gaps: 1
US-09-980-881A-4 (1-246) x CB161965 (1-505)
Qy 78 LeuTrpPheIleGlyHis----- 83
Db 3 TTGTGGTTCATAGGCCCATATACTCAATTCTATGGGATAATAGGCAATATACCAATCTC 62
Qy 83 ----- 83
Db 63 CTGAGGCTTGTGGATTCTATGTTATGCCAGTGGTTAATGTGGTGGTTATGACTACTCA 122
Qy 84 -----AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnHisCysIle 100
Db 123 TGGAAAAAGAATCTAATGTGGAGAAAGACCGTTCTTTCTATGCGAACATCATTTGCATC 182

```

```

Qy 101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
|||||
Db 183 GGAACAGACCTGAATAGGAACCTTGTCTCCAAACACTGGTGTGAGGAGGTGCATCCAGT 242
|||||
Qy 121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
|||||
Db 243 TCCTCATGCTCGGAAACCTACTCTGTGACTTTATCTCTGAGTCAGAACCCAGAGTGAAGGCA 302
|||||
Qy 141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
|||||
Db 303 GTGGCTAGTTTCTTGACAAGAAATATCAACCAAGATTAAAGCATACATCATCGCATTC 362
|||||
Qy 161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
|||||
Db 363 TACTCCAGCATATAGTGTTCATATTCTATACACAGTAAGTAAGCAAGACCATGAG 422
|||||
Qy 181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
|||||
Db 423 GAACTGCTCTTAATAGCCAGTGAAGCAGTTTCGTGCTATTGAGAAAAATTAGTAAAAATACC 482
|||||
Qy 201 ArgTyrThrHisGlyHisGly 207
|||||
Db 483 AGGTATACACATGGCCCATGGC 503

```

Search completed: January 2, 2004, 02:37:40
Job time : 2189.08 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:00:56 ; Search time 42.8831 Seconds
(without alignments)
910.539 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQYHSLNEIYSWIE.....IKYFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1338	100.0	246	21	AA11459 Human brain carbox
2	1338	100.0	338	21	AA11458 Human brain carbox
3	1325	99.0	360	21	AA11457 Human brain carbox
4	1262	94.3	386	24	ABG72095 Human protease pro
5	1257	93.9	386	24	ABG72093 Human protease pro
6	1239.5	92.6	423	17	AA190293 Human plasma carbo
7	1239.5	92.6	423	18	AA1914733 Human plasma carbo
8	1239.5	92.6	423	20	AA1952270 Human plasma carbo
9	1234.5	92.3	423	14	AA136273 Human plasma carbo

10	816.5	61.0	211	21	AA158129
11	612.5	45.8	324	22	AA166562
12	610	45.6	307	17	AA100602
13	582	43.5	306	16	AA175132
14	582	43.5	306	22	AA104477
15	582	43.5	404	16	AA175131
16	587	42.4	329	17	AA106174
17	567	42.4	349	17	AA106173
18	567	42.4	415	17	AA106172
19	567	42.4	415	19	AA174476
20	567	42.4	417	24	AB199688
21	567	42.4	424	17	AA106175
22	560	41.9	349	17	AA106181
23	560	41.9	349	18	AA113749
24	559	41.8	349	17	AA106182
25	559	41.8	349	18	AA113750
26	556	41.6	349	18	AA113763
27	554	41.4	349	18	AA113752
28	554	41.4	349	18	AA113758
29	553	41.3	349	18	AA113751
30	552	41.3	349	18	AA113753
31	552	41.3	417	23	AA184325
32	552	41.3	667	24	AB197414
33	551	41.2	311	22	AA166563
34	551	41.2	349	18	AA113760
35	551	41.2	349	18	AA113761
36	551	41.2	349	18	AA113762
37	550.5	41.1	354	22	AA166560
38	550.5	41.1	354	22	AA166566
39	550.5	41.1	362	22	AA174882
40	550.5	41.1	374	22	AA166547
41	550.5	41.1	374	22	AA166565
42	550.5	41.1	374	23	AA182703
43	550.5	41.1	374	24	AB137888
44	550.5	41.1	374	24	AB174658
45	550.5	41.1	444	23	AA124240

ALIGNMENTS

RESULT 1
AA11459
ID AA11459 standard; Protein; 246 AA.
AC AA11459;

DT 01-MAR-2001 (first entry)

XX Human brain carboxypeptidase B protein SEQ ID NO 4.

XX Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
XX treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
XX cerebroprotective; anti-Alzheimers; neurotropic; neuroprotective;
XX hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
XX Down's syndrome; head trauma.

OS Homo sapiens.

XX WO200066717-A1.

XX 09-NOV-2000.

PD 01-MAY-2000; 2000WO-JP02878.

XX 30-APR-1999; 99JP-0125169.

XX (MATS/) MATSUMOTO A.

XX Matsumoto A;

XX WPI; 2000-687534/67.

XX

PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
XX disorders -
PS Disclosure; Page 74-75; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimer's,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 246 AA;
Query Match 100.0%; Score 1338; DB 21; Length 246;
Best Local Similarity 100.0%; Pred. No. 5.8e-134;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLYLVKVSQKQAKNAI 60
DB 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLYLVKVSQKQAKNAI 60
QY 61 WIDCGIHAREWISPAFCLEWFGHNWRKNSFYANNHCIGTDLNRFASKHWCCEGASS 120
DB 61 WIDCGIHAREWISPAFCLEWFGHNWRKNSFYANNHCIGTDLNRFASKHWCCEGASS 120
QY 121 SSCSETYCGLYPSEPEVKAVASFLRNINQIKAYISMHSYSQHIYPPYSYTRSKSDHE 180
DB 121 SSCSETYCGLYPSEPEVKAVASFLRNINQIKAYISMHSYSQHIYPPYSYTRSKSDHE 180
QY 181 ELSLVAEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFTSNPPVEKL 240
DB 181 ELSLVAEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFTSNPPVEKL 240
QY 241 LPLSLK 246
DB 241 LPLSLK 246
RESULT 2
AAB11458
ID AAB11458 standard; Protein; 338 AA.
XX
AC AAB11458;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX

DR WPI; 2000-687534/67.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
XX disorders -
XX
PS Disclosure; Page 71-73; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimer's,
CC neurotropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 338 AA;
Query Match 100.0%; Score 1338; DB 21; Length 338;
Best Local Similarity 100.0%; Pred. No. 9.5e-134;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLYLVKVSQKQAKNAI 60
DB 93 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLYLVKVSQKQAKNAI 152
QY 61 WIDCGIHAREWISPAFCLEWFGHNWRKNSFYANNHCIGTDLNRFASKHWCCEGASS 120
DB 153 WIDCGIHAREWISPAFCLEWFGHNWRKNSFYANNHCIGTDLNRFASKHWCCEGASS 212
QY 121 SSCSETYCGLYPSEPEVKAVASFLRNINQIKAYISMHSYSQHIYPPYSYTRSKSDHE 180
DB 213 SSCSETYCGLYPSEPEVKAVASFLRNINQIKAYISMHSYSQHIYPPYSYTRSKSDHE 272
QY 181 ELSLVAEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFTSNPPVEKL 240
DB 273 ELSLVAEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFTSNPPVEKL 332
QY 241 LPLSLK 246
DB 333 LPLSLK 338
RESULT 3
AAB11457
ID AAB11457 standard; Protein; 360 AA.
XX
AC AAB11457;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimer's; neurotropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX

PI Matsumoto A;
XX WPI; 2000-687534/67.
DR N-PSDB; AAC81962.
XX
XX Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
XX Claim 1; Page 68-71; 84pp; Japanese.
XX
XX This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC neurotropic, neuroprotective and hemostatic activity, and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
XX Sequence 360 AA;
PS
XX
Query Match 99.0%; Score 1325; DB 21; Length 360;
Best Local Similarity 98.8%; Pred. No. 2.5e-132;
Matches 243; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAQNAI 60
DB 115 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAQNAI 174
QY 61 WIDCGIHAREWISPAFLWFIHNRMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 120
DB 175 WIDCGIHAREWISPAFLWFIHNRMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 234
QY 121 SSCSETYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSQHIYFPYSYTRSKSDHE 180
DB 235 SSCSETYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSQHIYFPYSYTRSKSDHE 294
QY 181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFTSNPPVEKL 240
DB 295 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFTSNPPVEKL 354
QY 241 LPLSLK 246
DB -355 LPLSLK 360
RESULT 4
ABG72095
ID ABG72095 standard; Protein; 386 AA.
XX
XX AC ABG72095;
XX
XX 14-FEB-2003 (first entry)
XX
XX Human protease protein, variation #2.
DE
XX
XX Human; enzyme; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW single nucleotide polymorphism; SNP.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Key
FT Misc-difference 169
FT /note= "This residue is the minor variant of a single
FT nucleotide polymorphism. The major variant is an
FT Ala residue"
XX
XX US2002137179-A1.

XX PD
XX 26-SEP-2002.
XX
XX 21-MAR-2001; 2001US-0813133.
XX PF
XX 21-MAR-2001; 2001US-0813133.
XX PR
XX (GANW/) GAN W.
XX PA (KETC/) KETCHUM K A.
XX PA (DFRA/) DI FRANCESCO V.
XX PA (BEAS/) BEASLEY E M.
XX
XX Gan W, Ketchum KA, Di Francesco V, Beasley EM;
PI WPI; 2003-102387/09.
XX N-PSDB; ABX13670, ABX13671.
DR
XX New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
PT -
XX
XX Claim 1; Page -; 80pp; English.
PS
XX
XX The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for
CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorders. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC protein, variant #2, the gene for which is located on chromosome 13.
CC Note: This sequence is not shown in the specification but has been
CC created by the indexer from information given in figure 3.
XX
XX SQ Sequence 386 AA;
Query Match 94.3%; Score 1262; DB 24; Length 386;
Best Local Similarity 99.6%; Pred. No. 1.5e-125;
Matches 231; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAQNAI 60
DB 115 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAQNAI 174
QY 61 WIDCGIHAREWISPAFLWFIHNRMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 120
DB 175 WIDCGIHAREWISPAFLWFIHNRMRKNSFYANNHCIGTDLNRNFASKHWCCEGASS 234
QY 121 SSCSETYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSQHIYFPYSYTRSKSDHE 180
DB 235 SSCSETYCGLYPSEPEVKAVASFLRRNINQIKAYISMHSYSQHIYFPYSYTRSKSDHE 294
QY 181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 295 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGDDWIYDLGIKYSFT 346

```
RESULT 5
ABG72093
ID ABG72093 standard; Protein; 386 AA.
AC ABG72093;
XX
XX
XX 14-FEB-2003 (first entry)
XX
XX Human protease protein, variation #1.
XX
XX Human; enzyme; protease; proteolytic degradation; proteolysis;
XX proliferation; differentiation; signalling; therapeutic; gene therapy;
XX protein therapy; diagnosis; immune response; vaccine; inflammation;
XX cancer; arteriosclerosis; degenerative disorder; chromosome 13;
XX single nucleotide polymorphism; SNP.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Misc-difference 169
XX /note= "This residue is the major variant of a single
XX nucleotide polymorphism. The minor variant is an
XX Thr residue"
XX
XX US2002131719-A1.
XX
XX 26-SEP-2002.
XX
XX 21-MAR-2001; 2001US-0813133.
XX
XX 21-MAR-2001; 2001US-0813133.
XX
XX (GANW/) GAN W.
XX (KETC/) KETCHUM K A.
XX (DFRA/) DI FRANCESCO V.
XX (BEAS/) BEASLEY E M.
XX
XX Gan W, Ketchum KA, Di Francesco V, Beasley EM;
XX WPI; 2003-102387/09.
XX N-PSDB; ABX13670, ABX13671.
XX
XX New isolated human protease proteins, useful for developing therapeutic
XX or diagnostic compositions, particularly for developing human
XX therapeutic agents that modulate protease activity in cells or tissues
XX
XX Claim 1; Fig 2; 80pp; English.
XX
XX The invention discloses an isolated human protease peptide, its allelic
XX variant or orthologue. The proper functioning of the cell requires
XX careful control of the levels of important structural proteins, enzymes
XX and regulatory proteins. One of the ways the cell reduces the steady
XX state level of a particular protein is by proteolytic degradation.
XX Proteolysis can also be used to convert a pre or pro-protein in to an
XX active form. Proteases also regulate many different cell proliferation,
XX differentiation and signalling processes. The peptides and nucleic acid
XX molecules are useful in the development of human therapeutics (gene and
XX protein therapy) and diagnostic compositions. The peptides are also
XX useful for raising antibodies or eliciting an immune response (vaccine),
XX as a reagent (including the labeled reagent) in assays designed to
XX quantitatively determine levels of the protein (or its binding partner or
XX ligand) in biological fluids, or as markers for tissues in which the
XX corresponding protein is preferentially expressed and in methods for
XX identifying a modulator of the peptide or an agent that binds to the
XX peptide. The agents identified are useful for treating protease-related
XX conditions that are specific for the subfamily of proteases that the
XX peptide belongs to, particularly in cells and tissues that express the
XX protease, such as inflammation, cancer, arteriosclerosis and degenerative
XX disorders. The modulator of the peptide is also useful for treating a
XX disorder characterised by an absence of, inappropriate or unwanted
XX expression of the protein. The sequence presented is the human protease
```

```
CC protein, variant #1, the gene for which is located on chromosome 13.
XX
XX Sequence 386 AA;
XX
XX Query Match 93.9%; Score 1257; DB 24; Length 386;
XX Best Local Similarity 99.1%; Pred. No. 5e-125;
XX Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 ASASYEYQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 60
XX 115 ASASYEYQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQAANKAI 174
XX
XX 61 WIDCGIHAREWISPAFLWFIHNRMRKRSFYANNHCIGTDLNRNFASKHWCCEGASS 120
XX 175 WIDCGIHAREWISPAFLWFIHNRMRKRSFYANNHCIGTDLNRNFASKHWCCEGASS 234
XX
XX 121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHIYFPYSYTRSKSKDHE 180
XX 235 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHIYFPYSYTRSKSKDHE 294
XX
XX 181 ELSLVASEAVRAIEKTSKNTRYTHGSETLYLAPCGGDDWIYDLGIKYSFT 232
XX 295 ELSLVASEAVRAIEKTSKNTRYTHGSETLYLAPCGGDDWIYDLGIKYSFT 346
XX
XX
XX RESULT 6
XX AAR90293
XX ID AAR90293 standard; Protein; 423 AA.
XX
XX AC AAR90293;
XX
XX 25-MAR-2003 (updated)
XX 12-APR-1996 (first entry)
XX
XX Human plasma carboxypeptidase B.
XX
XX Plasma carboxypeptidase B; hPCPB; antibody; detection;
XX purification; plasminogen; affinity column.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Active-site 181
XX /note= "Residue suspected of being involved in
XX catalytic activity of enzyme."
XX
XX Binding-site 183
XX /note= "Residue suspected of being part of the
XX binding site of enzyme."
XX
XX Binding-site 256..257
XX /note= "Residues suspected of being part of the
XX binding site of enzyme."
XX
XX Active-site 310
XX /note= "Residue suspected of being involved in
XX catalytic activity of enzyme."
XX
XX Binding-site 312
XX /note= "Residue suspected of being part of the
XX binding site of enzyme."
XX
XX Active-site 369
XX /note= "Residue suspected of being involved in
XX catalytic activity of enzyme."
XX
XX Binding-site 370..371
XX /note= "Residues suspected of being part of the
XX binding site of enzyme."
XX
XX Binding-site 394
XX /note= "Residue suspected of being part of the
XX binding site of enzyme."
XX
XX US5474901-A.
XX
XX 12-DEC-1995.
XX
XX 19-JUL-1994; 94US-0277540.
XX
```

```

PR 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
XX
PA (GETH ) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI; 1996-039508/04.
XX N-PSDB; AAT11671.
XX
DR Antbody to human plasma carboxypeptidase B - useful for detecting
PT and purifying hPCPB for use in treating clotting disorders e.g.
PT haemophilia A
XX
PS Disclosure; Figure 4; 40pp; English.
XX
CC An antibody which specifically binds human plasma carboxypeptidase B
CC (hPCPB) and does not cross react with other carboxypeptidases is
CC useful for the detection of hPCPB in vitro. The antibody is also
CC used for purifying hPCPB from a sample. Purification comprises
CC passing a sample thought to contain hPCPB over either a column to
CC which antibody has been bound, or a plasminogen affinity column,
CC eluting the column and then recovering the fraction containing the
CC hPCPB.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 423 AA;
Query Match 92.6%; Score 1239.5; DB 17; Length 423;
Best Local Similarity 86.2%; Pred. No. 4.2e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 ASASYEYOYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGSKQTAKNAI 60
Db 115 ASASYEYOYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGSKQTAKNAI 174
QY 61 WIDCGIHAREWISPAFLWFIGH-----83
Db 175 WIDCGIHAREWISPAFLWFIGHITQFYGIIGQYTLRLVDFVMPVNVVDGYDSWK 234
QY 84 NRWRKRSFYANNHCITGDLNRNFASKHWCCEGASSSCSETYCGLYPSEPEVKAVAS 143
Db 235 NRWRKRSFYANNHCITGDLNRNFASKHWCCEGASSSCSETYCGLYPSEPEVKAVAS 294
QY 144 FLRRNIQIKAYISMHSYSHIYPPVSYTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 295 FLRRNIQIKAYISMHSYSHIYPPVSYTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 354
QY 204 HGHGSETLYLAPGGDDWIDYDLGIKYSFT 232
Db 355 HGHGSETLYLAPGGDDWIDYDLGIKYSFT 383
RESULT 7
AAW14733
ID AAW14733 standard; Protein; 423 AA.
XX
AC AAW14733;
XX
DT 25-MAR-2003 (updated)
DT 08-MAY-1997 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW plasma; plasminogen.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22

```

```

FT Protein /note= "Signal peptide"
FT 23..423 /note= "Mature PCPB"
FT Cleavage-site 114 /note= "Potential clip site for activation of PCPB
FT as a carboxypeptidase"
FT 181
FT Active-site /note= "Expected to be involved in catalytic activity"
FT Binding-site 183 /note= "Expected to be involved in substrate binding"
FT 184 /note= "Expected to be involved in catalytic activity"
FT Binding-site 256 /note= "Expected to be involved in substrate binding"
FT Binding-site 257 /note= "Expected to be involved in substrate binding"
FT 310 /note= "Expected to be involved in catalytic activity"
FT Binding-site 312 /note= "Expected to be involved in substrate binding"
FT 369 /note= "Expected to be involved in catalytic activity"
FT Binding-site 370 /note= "Expected to be involved in substrate
FT binding, thought to determine specificity
FT of PCPB as a carboxypeptidase B"
FT 371 /note= "Expected to be involved in substrate binding"
FT Binding-site 394 /note= "Expected to be involved in substrate binding"
FT US5593674-A.
FT 14-JAN-1997.
XX
XX 27-APR-1995; 95US-0430787.
XX
PR 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
PR 27-APR-1995; 95US-0430787.
XX
PA (GETH ) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI; 1997-099413/09.
DR N-PSDB; AAT62846.
XX
XX Using human plasma carboxypeptidase B in blood coagulation - is
XX functionally related to carboxypeptidase A and pancreas
XX carboxypeptidase B
XX
PS Example 2; Column 37-42; 39pp; English.
XX
CC This sequence represents human plasma carboxypeptidase B (PCPB) which
CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
CC purified from human plasma or by transformed cell culture by
CC extraction using plasminogen bound to a solid phase.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 423 AA;
Query Match 92.6%; Score 1239.5; DB 18; Length 423;
Best Local Similarity 86.2%; Pred. No. 4.2e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 ASASYEYOYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGSKQTAKNAI 60
Db 115 ASASYEYOYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGSKQTAKNAI 174

```

```
Qy 61 WIDCGIHAREWISPAFLWFIGH----- 83
Db 175 WIDCGIHAREWISPAFLWFIGHTQFYGIIGQYTNLLRLVDFYVMPVNVVDGYDSWKK 234
Qy 84 NRMWRKNRSFYANNHCIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPSESEPEVKAVAS 143
Db 235 NRMWRKNRSFYANNHCIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPSESEPEVKAVAS 294
Qy 144 FLRRNINOIKAYISMHSYSQHIIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 295 FLRRNINOIKAYISMHSYSQHIIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354
Qy 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
Db 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

RESULT 8
AAW92270
ID AAW92270 standard; Protein; 423 AA.
XX
AC AAW92270;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB). thr147.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT Peptide 1..22
FT Protein /note= "signal peptide"
FT /note= "mature PCPBthr147"
XX
PN WO9855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD ) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI; 1999-045800/04.
DR N-PSDB; AAV74302.
XX
XX Detecting new polymorphism of human plasma carboxypeptidase B -
XX comprises Alanine or Threonine at position 147 of protein by DNA or
XX protein analysis, useful to detect risk of thrombotic disease in
XX humans
XX
PS Disclosure; Page 25-26; 35pp; English.
XX
CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
```

```
XX
SQ Sequence 423 AA;
Query Match 92.6%; Score 1239.5; DB 20; Length 423;
Best Local Similarity 86.2%; Pred. No. 4.2e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
Qy 1 ASASYEYQVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPYLVKVGSGEQTKAKNAI 60
Db 115 ASASYEYQVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPYLVKVGSGEQTKAKNAI 174
Qy 61 WIDCGIHAREWISPAFLWFIGH----- 83
Db 175 WIDCGIHAREWISPAFLWFIGHTQFYGIIGQYTNLLRLVDFYVMPVNVVDGYDSWKK 234
Qy 84 NRMWRKNRSFYANNHCIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPSESEPEVKAVAS 143
Db 235 NRMWRKNRSFYANNHCIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPSESEPEVKAVAS 294
Qy 144 FLRRNINOIKAYISMHSYSQHIIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 295 FLRRNINOIKAYISMHSYSQHIIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354
Qy 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
Db 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

RESULT 9
AAR36273
ID AAR36273 standard; Protein; 423 AA.
XX
AC AAR36273;
XX
DT 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT Peptide 1..22
FT Cleavage-site /label= signal_peptide
FT /note= "putative site of cleavage by trypsin to
FT activate PCPB as a carboxypeptidase"
FT Active-site 181
FT /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 183
FT /note= "expected to be involved in substrate
FT binding"
FT Active-site 184
FT /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 256..257
FT /note= "expected to be involved in substrate
FT binding"
FT Active-site 310
FT /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 312
FT /note= "expected to be involved in substrate
FT binding"
FT Active-site 369
FT /note= "expected to be involved in catalytic
FT activity"
FT Binding-site 370..371
FT /note= "expected to be involved in substrate
FT binding; Asp 370 is the residue which
FT determines specificity of the PCPB as
```


Db 120 DRWKRTRSTAGSSCLGVDPNRNF-NAGWCEVGASRSPCSEYTCGPAPSEKETKALAD 178
QY 144 FURNINQIKAYISMHSYQHIVFPYSYTRSKSKDHLSLVAESAARAEIKTSKNRYT 203
Db 179 FIRNLSLTIKAVLTIHYSQWMLYPYSYDKLPENYEELNALVKAAGKEL-ATLHGTYT 237
QY 204 HGHGSETLVAPCGDDWIYDLGIKYSFT 232
Db 238 YGPGATTIYPAAGGSDWSYDQGIKYSFT 266

RESULT 13
AAR75132
ID AAR75132 standard; Protein; 306 AA.
XX
AC AAR75132;
DT 13-MAR-1996 (first entry)
XX
DE Porcine carboxypeptidase B.
XX
KW Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
KW human serum albumin; premating factor alpha; mating factor alpha;
KW proCBB.
XX
OS Sus scrofa.
XX
PN WO9514096-A1.
XX
PD 26-MAY-1995.
XX
PF 16-NOV-1994; 94WO-US13142.
XX
PR 16-NOV-1993; 93US-0153258.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
PI Sterner JL, Zhang H;
XX
DR WPI: 1995-200386/26.
DR N-PSDB; AAQ90601.
XX
PT DNA encoding porcine carboxypeptidase B - used for transforming
PT host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS Example 1; Page 18; 34pp; English.
XX
CC The porcine carboxypeptidase B coding sequence can be place in a
CC bacterial or pref. Pichia yeast expression vector. The expression
CC vector further comprises the signal peptide of either human serum
CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC (designated pLGD27 - NRRL B-21027). The method can be used for
CC producing large amounts of porcine carboxypeptidase B and when
CC produced in Pichia yeast, the protein does not need solubilisation
CC or folding. The produced enzyme is then used for pref. cleaving
CC basic residues from the carboxy terminus of proteins.
XX
SQ Sequence 306 AA;

Query Match 43.5%; Score 582; DB 16; Length 306;
Best Local Similarity 44.3%; Pred. No. 3.2e-53;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIFITERHPDMLTKIHIGSSFYKYPVLVKSQKEQTAKNAIWDCG 65
Db 6 YEKYNWETIEAWTKQVTSNPDLISRTAIGTTFLGNVYLLKV-GKPGNKPAIFWDCG 64
QY 66 IHAREWISPAFCWFI-----GH-----NRMR 88
Db 65 FHAREWISPAFCWFI-----GH-----NRMR 124

QY 89 KQSFYANNHCIGTDLNRNFASKHCEGASSSSCSEYTCGLYPESEPEVKAVASFLRN 148
Db 125 KTRSTNAGTTCIGTDPNRF-DAGRCTTGASTDPCDETCGSAESEKETKALADIRN 183
QY 149 INQIKAYISMHSYQHIVFPYSYTRSKSKDHLSLVAESAARAEIKTSKNRYTHGHS 208
Db 184 LSSIKAYLTIHYSQWMLYPYSYDKLPENNAELNNAKAAVKEL-ATLYGTYKTYGPA 242
QY 209 ETLYLAPCGDDWIYDLGIKYSFT 232
Db 243 TTIYPAAGGSDWDWAYDQGIKYSFT 266

RESULT 14
AAU04477
ID AAU04477 standard; Protein; 306 AA.
XX
AC AAU04477;
DT 26-SEP-2001 (first entry)
XX
DE Porcine carboxypeptidase B (CpB) protein.
XX
KW Carboxypeptidase B; CpB; pig; protease activity; trypsin; chymotrypsin;
KW animal product; purification; contaminant; biopharmaceutical agent;
KW zinc-containing pancreatic exopeptidase; insulin; proteolytic processing;
KW protein sequence analysis.
XX
OS Sus scrofa.
XX
PN WO200151624-A2.
XX
PD 19-JUL-2001.
XX
PF 12-JAN-2001; 2001WO-US000008.
XX
PR 12-JAN-2000; 2000US-0175781.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Hanquier JM;
XX
DR WPI; 2001-442143/47.
XX
PT New carboxypeptidase B, which is free of animal products and
PT contaminating enzyme activity, useful for producing biopharmaceutical
PT agents (e.g. insulin) and biologically active polypeptides, as well as
PT in protein sequence analysis -
XX
PS Claim 4; Page 41-42; 42pp; English.
XX
CC The sequence represents porcine carboxypeptidase B (CpB), which is
CC essentially free of contaminating protease activity (e.g. trypsin or
CC chymotrypsin activity) and free of animal products, bacterial cell
CC components and other contaminants that result from purification
CC processes. CpB is a zinc-containing pancreatic exopeptidase which
CC specifically removes C-terminal arginine, lysine and ornithine from
CC peptides. The CpB of the invention is useful for commercial and research
CC purposes, e.g. in production of biopharmaceutical agents, such as insulin
CC and other biologically active polypeptides, as well as in protein
CC sequence analysis. Compositions containing CpB are useful in preparing
CC proteins which are in need of proteolytic processing in order to form the
CC mature protein.
XX
SQ Sequence 306 AA;

Query Match 43.5%; Score 582; DB 22; Length 306;
Best Local Similarity 44.3%; Pred. No. 3.2e-53;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIFITERHPDMLTKIHIGSSFYKYPVLVKSQKEQTAKNAIWDCG 65
Db 6 YEKYNWETIEAWTKQVTSNPDLISRTAIGTTFLGNVYLLKV-GKPGNKPAIFWDCG 64
QY 66 IHAREWISPAFCWFI-----GH-----NRMR 88
Db 65 FHAREWISPAFCWFI-----GH-----NRMR 124

Db 6 YEKNNWETIEAWTKQVTSNPDLISRTAIGTTFLGNIIYLLKV-GKPGNKPDAIFMDCG 64
 QY 66 IHAREWISPAFCLWFI-:-----GH-----NRMR 88
 Db 65 FHAREWISHAFQCFVREAVLTGYESHMTFELINKLDFVLPVNLIDGYIYTWTKRMR 124
 QY 89 KNRSFYANNHCIGTDLNRNPFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRN 148
 Db 125 KTRSTNAGTTCIGTDPNRF-DAGWCTTGASTPCDETYCGSAASEKETKALADFIRN 193
 QY 149 INQIKAYISMHSYQSHIVFPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
 Db 184 LSSIKAYLTTHSYQSHILYPSYDYKLPENNAELNNAKAAVKEL-ATLYGTYTYGPGA 242
 QY 209 ETLYLAPGGDDWYDLGIKYSFT 232
 Db 243 TTIYPAAGSGDDWAYDQGIKYSFT 266

RESULT 15

AAR75131 ID AAR75131 standard; Protein; 404 AA.

XX AC AAR75131;

DT 13-MAR-1996 (first entry)

XX Porcine Tyr-His-Met Procarboxypeptidase B.

XX Procarboxypeptidase B; carboxypeptidase B; Pichia; pCPB;
 KW human serum albumin; premating factor alpha; mating factor alpha;
 KW proCBB.

XX Sus scrofa.

XX WO9514096-A1.

XX 26-MAY-1995.

XX 16-NOV-1994; 94WO-US13142.

XX 16-NOV-1993; 93US-0153258.

XX (ELIL) LILLY & CO ELI.

XX Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
 PI Sterner JL, Zhang H;

XX WPI; 1995-200386/26.

XX N-PSDB; AAQ90600.

XX DNA encoding porcine carboxypeptidase B - used for transforming
 PT host cells, partic. Pichia species, for prodn. of the enzyme

XX Example 1; Page 16-17; 34pp; English.

XX The porcine carboxypeptidase B coding sequence can be place in a
 CC bacterial or pref. Pichia yeast expression vector. The expression
 CC vector further comprises the signal peptide of either human serum
 CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
 CC (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
 CC pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
 CC (designated pLGD27 - NRRL B-21027). The method can be used for
 CC producing large amounts of porcine carboxypeptidase B and when
 CC produced in Pichia yeast, the protein does not need solubilisation
 CC or folding. The produced enzyme is then used for pref. cleaving
 CC basic residues from the carboxy terminus of proteins.

XX Sequence 404 AA;

Query Match 43.5%; Score 582; DB 16; Length 404;
 Best Local Similarity 44.3%; Pred. No. 4.9e-53;
 Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

QY 6 YEOYHSLNEIYSWIEPITERHPDMLTKIHIGSSFYKVPYLYVLKVSKEQTAKVAIWIDCG 65
 Db 104 YEKNNWETIEAWTKQVTSNPDLISRTAIGTTFLGNIIYLLKV-GKPGNKPDAIFMDCG 162
 QY 66 IHAREWISPAFCLWFI-:-----GH-----NRMR 88
 Db 163 FHAREWISHAFQCFVREAVLTGYESHMTFELINKLDFVLPVNLIDGYIYTWTKRMR 222
 QY 89 KNRSFYANNHCIGTDLNRNPFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRN 148
 Db 223 KTRSTNAGTTCIGTDPNRF-DAGWCTTGASTPCDETYCGSAASEKETKALADFIRN 281
 QY 149 INQIKAYISMHSYQSHIVFPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
 Db 282 LSSIKAYLTTHSYQSHILYPSYDYKLPENNAELNNAKAAVKEL-ATLYGTYTYGPGA 340
 QY 209 ETLYLAPGGDDWYDLGIKYSFT 232
 Db 341 TTIYPAAGSGDDWAYDQGIKYSFT 364

Search completed: January 1, 2004, 19:28:08
 Job time : 44.8831 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:21:46 ; Search time 15.1503 Seconds
(without alignments)
687.014 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQYHSLNEIYSWIE.....IKVSTSNPPVKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata2/iaa/5B COMB.pep.*
3: /cgn2_6/prodata2/iaa/6A COMB.pep.*
4: /cgn2_6/prodata2/iaa/6B COMB.pep.*
5: /cgn2_6/prodata2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1257	93.9	386	4	US-09-813-133A-2
2	1239.5	92.6	423	1	US-07-649-591B-3
3	1239.5	92.6	423	1	US-08-277-540-3
4	1239.5	92.6	423	1	US-08-430-787A-3
5	1239.5	92.6	423	2	US-08-869-057-2
6	1228.5	91.8	423	4	US-09-813-133A-4
7	610	45.6	307	2	US-08-782-760-6
8	610	45.6	307	5	PCT-US96-00995-6
9	582	43.5	306	1	US-08-696-139-4
10	582	43.5	404	1	US-08-696-139-2
11	567	42.4	329	4	US-09-011-769A-51
12	567	42.4	349	4	US-09-011-769A-47
13	567	42.4	415	2	US-08-860-882A-57
14	567	42.4	415	4	US-09-011-769A-39
15	567	42.4	424	4	US-09-011-769A-56
16	561	41.9	417	1	US-07-649-591B-7
17	561	41.9	417	1	US-08-277-540-7
18	561	41.9	417	1	US-08-430-787A-7
19	560	41.9	349	4	US-09-011-769A-60
20	559	41.8	349	4	US-09-011-769A-64
21	554	41.4	396	1	US-07-649-591B-4
22	554	41.4	396	1	US-08-277-540-4
23	554	41.4	396	1	US-08-430-787A-4
24	552	41.3	417	1	US-07-649-591B-6
25	552	41.3	417	1	US-08-277-540-6
26	552	41.3	417	1	US-08-430-787A-6
27	549	41.0	437	4	US-09-675-305-10

28	549	41.0	613	3	US-09-171-945-113	Sequence 113, App
29	549	41.0	716	3	US-09-171-945-125	Sequence 125, App
30	492	36.8	399	4	US-09-710-099-8	Sequence 8, Appli
31	487.5	36.4	419	1	US-07-649-591B-5	Sequence 5, Appli
32	487.5	36.4	419	1	US-08-277-540-5	Sequence 5, Appli
33	487.5	36.4	419	1	US-08-430-787A-5	Sequence 5, Appli
34	483	36.1	417	1	US-07-649-591B-8	Sequence 8, Appli
35	483	36.1	417	1	US-08-277-540-8	Sequence 8, Appli
36	483	36.1	417	1	US-08-430-787A-8	Sequence 8, Appli
37	471.5	35.2	350	4	US-09-675-305-12	Sequence 12, Appli
38	465.5	34.8	419	3	US-08-640-906-2	Sequence 2, Appli
39	465.5	34.8	419	4	US-09-395-936-2	Sequence 2, Appli
40	465.5	34.8	436	4	US-09-710-099-6	Sequence 6, Appli
41	460.5	34.4	419	3	US-08-640-906-17	Sequence 17, Appli
42	460.5	34.4	419	4	US-09-395-936-17	Sequence 17, Appli
43	459	34.3	417	3	US-08-640-906-4	Sequence 4, Appli
44	459	34.3	417	4	US-09-395-936-4	Sequence 4, Appli
45	458	34.2	417	3	US-08-640-906-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Query Match	93.9%	Score 1257;	DB 4;	Length 386;
Best Local Similarity	99.1%	Pred. No. 9.6e-129;		
Matches	230;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
QY	1	ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPYLYLVKVSQKQAKNAI	60	
DB	115	ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPYLYLVKVSQKQAKNAI	174	
QY	61	WIDCGIHAREWISPAFLWFIHGNRMWRKNRNFYANNHCIGTDLNRNFASKHWCCEGASS	120	
DB	175	WIDCGIHAREWISPAFLWFIHGNRMWRKNRNFYANNHCIGTDLNRNFASKHWCCEGASS	234	
QY	121	SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSYTRSKSDHE	180	
DB	235	SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSHQIVFPYSYTRSKSDHE	294	
QY	181	ELSLVASEAVRAIEKTSKNTRYTHGSGSTLYLAPCGGDDWIYDLGIKYSFT	232	
DB	295	ELSLVASEAVRAIEKTSKNTRYTHGSGSTLYLAPCGGDDWIYDLGIKYSFT	346	
RESULT 2				
US-07-649-591B-3				
; Sequence 3, Application US/07649591B				
; Patent No. 5206161				
; GENERAL INFORMATION:				
; APPLICANT: Dennis Drayna and Daniel Eaton				
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase				
; NUMBER OF SEQUENCES: 8				
; CORRESPONDENCE ADDRESS:				

ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/649,591B
FILING DATE: 19910201
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/266-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear

US-07-649-591B-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 8.9e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 60
DB 115 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 174

QY 61 WIDCGIHAREWISPAFLWFIGH----- 83
DB 175 WIDCGIHAREWISPAFLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVNVVDGYDSWK 234

QY 84 NRMWRKRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVAS 143
DB 235 NRMWRKRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVAS 294

QY 144 FLRRNINQKAYISMHSYSHQHVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 295 FLRRNINQKAYISMHSYSHQHVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

RESULT 3
US-08-277-540-3
Sequence 3, Application US/08277540
Patent No. 5474901

GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,540
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992

PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-08-277-540-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 8.9e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 60
DB 115 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYLVKSGKEQTAKNAI 174

QY 61 WIDCGIHAREWISPAFLWFIGH----- 83
DB 175 WIDCGIHAREWISPAFLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVNVVDGYDSWK 234

QY 84 NRMWRKRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVAS 143
DB 235 NRMWRKRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVAS 294

QY 144 FLRRNINQKAYISMHSYSHQHVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 295 FLRRNINQKAYISMHSYSHQHVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

RESULT 4
US-08-430-787A-3
Sequence 3, Application US/08430787A
Patent No. 5593674

GENERAL INFORMATION:
APPLICANT: Drayna, Dennis T., Eaton, Dan L.
TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,787A
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277,540
FILING DATE: 19-JUL-1994
APPLICATION NUMBER: 08/167727
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/959944
FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/649591
FILING DATE: 01-FEB-91
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 689D1C1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-430-787A-3

Query Match 92.6%; Score 1239.5; DB 1; Length 423;
Best Local Similarity 86.2%; Pred. No. 8.9e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKNAI 60
DB 115 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKNAI 174
QY 61 WDCGTHAREWISPAFLWFIH----- 83
DB 175 WDCGTHAREWISPAFLWFIHITQFYGIQYTNLLRLVDFYVMPVNVVDGYDYSWKK 234
QY 84 NRMWRNRSFYANNHCIGTDLNRFASKHWCCEGASSSSCSEYCYGLYPESEPEVKAVAS 143
DB 235 NRMWRNRSFYANNHCIGTDLNRFASKHWCCEGASSSSCSEYCYGLYPESEPEVKAVAS 294
QY 144 FLRRNINQIKAYISMHSYSHQHVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 295 FLRRNINQIKAYISMHSYSHQHVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

RESULT 5
US-08-869-057-2
Sequence 2, Application US/08869057
Patent No. 5985562
GENERAL INFORMATION:
APPLICANT: Morser, Michael J
APPLICANT: Nagashima, Mariko
TITLE OF INVENTION: Method of Detecting Thrombotic Disease
TITLE OF INVENTION: Risk
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Berlex Biosciences Legal Department
STREET: 15049 San Pablo Avenue

CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94804-0099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,057
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Washtien, Wendy L
REGISTRATION NUMBER: 36,301
REFERENCE/DOCKET NUMBER: 51509AUSM1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-262-5411
TELEFAX: 510-262-7095
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
TISSUE TYPE: Plasma
FEATURE:
NAME/KEY: Peptide
LOCATION: 23..401
US-08-869-057-2

Query Match 92.6%; Score 1239.5; DB 2; Length 423;
Best Local Similarity 86.2%; Pred. No. 8.9e-127;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
QY 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKNAI 60
DB 115 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYLYLVKVGKEQTAKNAI 174
QY 61 WDCGTHAREWISPAFLWFIH----- 83
DB 175 WDCGTHAREWISPAFLWFIHITQFYGIQYTNLLRLVDFYVMPVNVVDGYDYSWKK 234
QY 84 NRMWRNRSFYANNHCIGTDLNRFASKHWCCEGASSSSCSEYCYGLYPESEPEVKAVAS 143
DB 235 NRMWRNRSFYANNHCIGTDLNRFASKHWCCEGASSSSCSEYCYGLYPESEPEVKAVAS 294
QY 144 FLRRNINQIKAYISMHSYSHQHVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 295 FLRRNINQIKAYISMHSYSHQHVFPYSTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 HGHGSETLYLAPGGDDWIYDLGIKYSFT 383

RESULT 6
US-09-813-133A-4
Sequence 4, Application US/09813133A
Patent No. 6455294
GENERAL INFORMATION:
APPLICANT: GAN, Weiniu et al
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: CL001173
CURRENT APPLICATION NUMBER: US/09/813,133A
CURRENT FILING DATE: 2001-06-06
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0

Db 1 ASGSHYKNNWETIEAWIQOAVDPDLVTSQVIGTTEGRNMYLKI-GKTRPNKPAI 59
QY 61 WIDCGHAREWISPAFLWFI-----GH----- 83
Db 60 FIDCGFHAREWISPAFLWFI-----GH----- 119
QY 84 NRMWRNRSFYANNHCIGTDLNRNFASKHWCCEGASSSSCSEYCGLYPESEPEVKAVAS 143
Db 120 DRMRKTRSTMACSSCLGVDPNRF-NAGMCEVGASRSPCSEYCCGPAPSEKETKALAD 178
QY 144 FLARNINQKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 179 FIRNNLSIKAYLTIHSYQMLPYSDYKLPENNELNALVKGAKEL-ATLHGTKYT 237
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
Db 238 YGPGATTIYPAAGGSDWSYDQGIKYSFT 266

RESULT 9
US-08-696-139-4
; Sequence 4, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Herzhberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-696-139-4

Query Match 43.5%; Score 582; DB 1; Length 306;
Best Local Similarity 44.3%; Pred. No. 3-2e-55;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKYPYLVKVGSGEQAKNAIWDG 65

Db 6 YEKYNWETIEAWTQVTSNPDLISRTAIGTTFGLNNIYLLKV-GKPGPNKPAI FMDCG 64
QY 66 IHAREWISPAFLWFI-----GH-----NRMR 88
Db 65 FHAREWISPAFLWFI-----GH-----NRMR 124
QY 89 KRSFYANNHCIGTDLNRNFASKHWCCEGASSSSCSEYCGLYPESEPEVKAVASFLRN 148
Db 125 KTRSTNAGTTCGTDPNRF-NAGMCEVGASRSPCSEYCCGPAPSEKETKALADFIRN 183
QY 149 INQIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db 184 LSIKAYLTIHSYQMLPYSDYKLPENNELNALVKGAKEL-ATLYGKTYTGP 242
QY 209 ETLYLAPGGDDWIYDLGIKYSFT 232
Db 243 TTIYPAAGGSDWDWAYDQGIKYSFT 266

RESULT 10
US-08-696-139-2
; Sequence 2, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Herzhberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-696-139-2

Query Match 43.5%; Score 582; DB 1; Length 404;
Best Local Similarity 44.3%; Pred. No. 4.8e-55;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFYKYPYLVKVGSGEQAKNAIWDG 65

Db 104 YKYNWETIEAWTKVTSNPOLISRTAIGTFLGNVYLLKV-GKPGNKPAIFMDCG 162
Qy 66 IHAREWISPAFLWFI-----GH-----NMWR 88
Db 163 FHAREWISPAFLWFI-----GH-----NMWR 222
Qy 89 KNRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESPEVKAVASFLRRN 148
Db 223 KTRSTNAGTTCIGTDPNRF-DAGWCTTGASTPDCDETYCGSAAESKETKALADFIRN 281
Qy 149 INQKAYISMHSYQHIVFPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTYTHGHS 208
Db 282 LSSIKAYLTHSYQMLYPYSYDYKLPENNAELNLAKAAVKEL-ATLGTGYTYPGA 340
Qy 209 ETLYPAGGGDDWIYDLGIKYSFT 232
Db 341 TTIYPAGGGDDWAYDQIRYSFT 364

RESULT 11

US-09-011-769A-51

; Sequence 51, Application US/09011769A

; Patent No. 6436691

; GENERAL INFORMATION:

; APPLICANT: SLATER, Anthony M.

; BLAKEY, David C.

; DAVIES, David H.

; HENNAM, John F.

; HENNEQUIN, Laurent F.A.

; MARSHAM, Peter R.

; DOWELL, Robert I.

; TITLE OF INVENTION: Chemical Compounds

; NUMBER OF SEQUENCES: 87

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pillsbury Madison & Sutro, LLP

; STREET: 1100 New York Ave., N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 1.44 Mb disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MS word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/011,769A

; FILING DATE: 13-Feb-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB96/01975

; FILING DATE: 13-AUG-1996

; APPLICATION NUMBER: GB 9612295.7

; FILING DATE: 12-JUN-1996

; APPLICATION NUMBER: GB 9611019.2

; FILING DATE: 25-MAY-1996

; APPLICATION NUMBER: GB 9516810.0

; FILING DATE: 16-AUG-1995

; INFORMATION FOR SEQ ID NO: 51:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 329 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 51:

US-09-011-769A-51

Query Match 42.4%; Score 567; DB 4; Length 329;

Best Local Similarity 41.3%; Pred. No. 1.5e-53;

Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

Qy 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPYLVLVKVSKEQTAKNAI 60

Db 23 ATCHSEYKYNKWTIEAWTQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 81
Qy 61 WIDCGIHAREWISPAFLWFI-----GH-----NMWR 83
Db 82 FMDCGFHAREWISPAFLWFI-----GH-----NMWR 141
Qy 84 NRMWRKNSRFPYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESPEVKAVAS 143
Db 142 SRWRKTRSTHTGSSCIGTDPNRF-DAGWCEIGASRNPCDETYCGSAAESKETKALAD 200
Qy 144 FLERNINQIKAYISMHSYQHIVFPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTYTH 203
Db 201 FIENKLSSIKAYLTHSYQMLYPYSYDYKLPENNAELNLAKAAVKEL-ASLHGTYT 259
Qy 204 HGHSETLYLAPGGDDWIYDLGIKYSFT 232
Db 260 YGPGATTIYPAGGGDDWAYDQIRYSFT 288

RESULT 12

US-09-011-769A-47

; Sequence 47, Application US/09011769A

; Patent No. 6436691

; GENERAL INFORMATION:

; APPLICANT: SLATER, Anthony M.

; BLAKEY, David C.

; DAVIES, David H.

; HENNAM, John F.

; HENNEQUIN, Laurent F.A.

; MARSHAM, Peter R.

; DOWELL, Robert I.

; TITLE OF INVENTION: Chemical Compounds

; NUMBER OF SEQUENCES: 87

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pillsbury Madison & Sutro, LLP

; STREET: 1100 New York Ave., N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 1.44 Mb disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MS word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/011,769A

; FILING DATE: 13-Feb-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB96/01975

; FILING DATE: 13-AUG-1996

; APPLICATION NUMBER: GB 9612295.7

; FILING DATE: 12-JUN-1996

; APPLICATION NUMBER: GB 9611019.2

; FILING DATE: 25-MAY-1996

; APPLICATION NUMBER: GB 9516810.0

; FILING DATE: 16-AUG-1995

; INFORMATION FOR SEQ ID NO: 47:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 349 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-011-769A-47

Query Match 42.4%; Score 567; DB 4; Length 349;

Best Local Similarity 41.3%; Pred. No. 1.7e-53;

Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

Qy 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPYLVLVKVSKEQTAKNAI 60

Db 23 ATGSHSEKYNKWTETIAWTOQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 81
QY 61 WIDCGIHAREWISPAFLWFI-----GH----- 83
Db 82 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFVLPVLNIDGYIYTWK 141
QY 84 NRMWRKRSFYANNHCIGTDLNRNFASKHWCSEAGSSSSCSEYTCGLYPESEPEVKAVAS 143
Db 142 SRFWKTRSTHTGSSCIGTDPNRF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 200
QY 144 FLRRNINOIKAYISMHSYOHIVFPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTVT 203
Db 201 FIRKLSSIKAYLTTHSYQMMIYPSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 259
QY 204 HGGSETLYLAPGGDDWIYDLGIRYSFT 232
Db 260 YGPGATTIYPAAGSGDDWAYDQGIYSFT 288

RESULT 13

US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGLETE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM, JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-860-882A-57

Query Match 42.4%; Score 567; DB 2; Length 415;
Best Local Similarity 41.3%; Pred. No. 2.2e-53;
Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

QY 1 ASASYEQHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPLVVLKVSKEQAKNAI 60
Db 109 ATGSHSEKYNKWTETIAWTOQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 167
QY 61 WIDCGIHAREWISPAFLWFI-----GH----- 83
Db 168 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFVLPVLNIDGYIYTWK 227
QY 84 NRMWRKRSFYANNHCIGTDLNRNFASKHWCSEAGSSSSCSEYTCGLYPESEPEVKAVAS 143
Db 228 SRFWKTRSTHTGSSCIGTDPNRF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 286
QY 144 FLRRNINOIKAYISMHSYOHIVFPYSYTRSKSDHEELSLVASEAVRAIEKTSKNTVT 203
Db 287 FIRKLSSIKAYLTTHSYQMMIYPSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 345
QY 204 HGGSETLYLAPGGDDWIYDLGIRYSFT 232
Db 346 YGPGATTIYPAAGSGDDWAYDQGIYSFT 374

RESULT 14

US-09-011-769A-39
; Sequence 39, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; APPLICANT: BLAKEY, David C.
; APPLICANT: DAVIES, David H.
; APPLICANT: HENNAM, John F.
; APPLICANT: HENNEQUIN, Laurent F.A.
; APPLICANT: MARSHAM, Peter R.
; APPLICANT: DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011,769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-011-769A-39

Query Match 42.4%; Score 567; DB 4; Length 415;
Best Local Similarity 41.3%; Pred. No. 2.2e-53;
Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

QY 1 ASASYEYQVHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPLVYLKVSKEQAKNAI 60
Db 109 ATGHSYKYNKWTETIAWTOQVATENPALISRSVIGTFEGRAIYLLKV-GKAGQNKPAI 167
QY 61 WIDCGIHAREWISPAFCLWFI-----GH----- 83
Db 168 FMDCGFHAREWISPAFCQFVREAVRTYGREIQVTELLDKLDFVLPVNLIDGYIYTWK 227
QY 84 NRMWRKRSFYANNHCIGTDLNRNPFASKHWCBEAGSSSCSEYTCGLYPESEBEVKAVAS 143
Db 228 SRFWRKTRSTHTGSSCIGTDPRNF-DAGWCEIGASRNPCEYTCGPAASEKETKALAD 286
QY 144 FLARNINOIKAYISMHSYSOHIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 287 FIRKLUSSIKAYLTHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 345
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
Db 346 YGPGATTIYPAGGSDWAYDQIRYSFT 374

RESULT 15

US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:

APPLICANT: SLATER, Anthony M.
BLAKEY, David C.
DAVIES, David H.
HENNAM, John F.
HENNEQUIN, Laurent F. A.
MARSHAM, Peter R.
DOWELL, Robert I.
ADDRESSEE: Pillebury Madison & Sutro, LLP
CORRESPONDENCE ADDRESS:
STREET: 1100 New York Ave., N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 Mb disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/011,769A
FILING DATE: 13-Feb-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/01975
FILING DATE: 13-AUG-1996
APPLICATION NUMBER: GB 9612295.7
FILING DATE: 12-JUN-1996
APPLICATION NUMBER: GB 9611019.2
FILING DATE: 25-MAY-1996
APPLICATION NUMBER: GB 9516810.0
FILING DATE: 16-AUG-1995

INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 56:

US-09-011-769A-56

Query Match 42.4%; Score 567; DB 4; Length 424;
Best Local Similarity 41.3%; Pred. No. 2.3e-53;
Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

QY 1 ASASYEYQVHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPLVYLKVSKEQAKNAI 60
Db 118 ATGHSYKYNKWTETIAWTOQVATENPALISRSVIGTFEGRAIYLLKV-GKAGQNKPAI 176
QY 61 WIDCGIHAREWISPAFCLWFI-----GH----- 83
Db 177 FMDCGFHAREWISPAFCQFVREAVRTYGREIQVTELLDKLDFVLPVNLIDGYIYTWK 236
QY 84 NRMWRKRSFYANNHCIGTDLNRNPFASKHWCBEAGSSSCSEYTCGLYPESEBEVKAVAS 143
Db 237 SRFWRKTRSTHTGSSCIGTDPRNF-DAGWCEIGASRNPCEYTCGPAASEKETKALAD 295
QY 144 FLARNINOIKAYISMHSYSOHIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 296 FIRKLUSSIKAYLTHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 354
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
Db 355 YGPGATTIYPAGGSDWAYDQIRYSFT 383

Search completed: January 1, 2004, 19:33:11
Job time : 16.1503 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:32:17 ; Search time 35.1795 Seconds
(without alignments)
1403.024 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQVHSLNEIYSWIE.....IKYFTSNPPVVKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues
Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1239.5	92.6	423	12	US-10-379-836-17
2	1164.5	87.0	423	12	US-10-379-836-2
3	1105.5	82.6	422	12	US-10-379-836-18
4	1095.5	81.9	422	12	US-10-379-836-16
5	816.5	61.0	211	9	US-09-925-102-467
6	567	42.4	402	12	US-10-379-836-20
7	552	41.3	417	12	US-10-341-434-188
8	552	41.3	417	15	US-10-229-546-2
9	552	41.3	417	15	US-10-229-546-9
10	550.5	41.1	374	9	US-09-888-615-61
11	550.5	41.1	444	16	US-10-176-306-74
12	549	41.0	437	12	US-10-274-639-12
13	549	41.0	437	14	US-10-200-344-10
14	549	41.0	613	10	US-09-910-059-113
15	549	41.0	716	10	US-09-910-059-125

16	501.5	37.5	298	12	US-10-379-836-19	Sequence 19, Appl
17	492	36.8	399	15	US-10-200-910-8	Sequence 8, Appl
18	471.5	35.2	350	14	US-10-200-344-12	Sequence 12, Appl
19	465.5	34.8	406	9	US-09-925-297-517	Sequence 517, Appl
20	465.5	34.8	419	12	US-10-345-680-65	Sequence 65, Appl
21	465.5	34.8	436	15	US-10-200-910-6	Sequence 6, Appl
22	458	34.2	421	11	US-09-946-374-234	Sequence 234, Appl
23	458	34.2	421	12	US-10-015-387A-234	Sequence 234, Appl
24	458	34.2	421	12	US-10-006-130A-234	Sequence 234, Appl
25	458	34.2	421	12	US-10-199-672-308	Sequence 308, Appl
26	458	34.2	421	12	US-10-006-172A-234	Sequence 234, Appl
27	458	34.2	421	12	US-10-187-749-308	Sequence 308, Appl
28	458	34.2	421	12	US-10-194-457-308	Sequence 308, Appl
29	458	34.2	421	12	US-10-184-642-308	Sequence 308, Appl
30	458	34.2	421	12	US-10-196-747-308	Sequence 308, Appl
31	458	34.2	421	12	US-10-015-392A-234	Sequence 234, Appl
32	458	34.2	421	12	US-10-017-253A-234	Sequence 234, Appl
33	458	34.2	421	12	US-10-173-689-308	Sequence 308, Appl
34	458	34.2	421	12	US-10-173-690-308	Sequence 308, Appl
35	458	34.2	421	12	US-10-173-691-308	Sequence 308, Appl
36	458	34.2	421	12	US-10-173-692-308	Sequence 308, Appl
37	458	34.2	421	12	US-10-173-694-308	Sequence 308, Appl
38	458	34.2	421	12	US-10-173-698-308	Sequence 308, Appl
39	458	34.2	421	12	US-10-173-699-308	Sequence 308, Appl
40	458	34.2	421	12	US-10-173-707-308	Sequence 308, Appl
41	458	34.2	421	12	US-10-174-569-308	Sequence 308, Appl
42	458	34.2	421	12	US-10-174-583-308	Sequence 308, Appl
43	458	34.2	421	12	US-10-174-587-308	Sequence 308, Appl
44	458	34.2	421	12	US-10-174-589-308	Sequence 308, Appl
45	458	34.2	421	12	US-10-174-591-308	Sequence 308, Appl

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; TYPE: PPT
; LENGTH: 423
; ORGANISM: Homo sapiens
US-10-379-836-17

Query Match	92.6%;	Score	1239.5;	DB	12;	Length	423;		
Best Local Similarity	86.2%;	Pred. No.	2.5e-121;						
Matches	232;	Conservative	0;	Mismatches	0;	Indels	37;	Gaps	1;
Qy	1	ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHIGSSF	EKYPYLVKVSQKEQTAKNAI	60					
Db	115	ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHIGSSF	EKYPYLVKVSQKEQTAKNAI	174					
Qy	61	WDCGHAREWISPAFLWFI	GH-----	83					
Db	175	WDCGHAREWISPAFLWFI	GHITQFYGIQYTNLLRLVDFYVNVVGVGYDYSWK	234					
Qy	84	NRMWRKRSFYANNHCIGTDLNRNFASKKWCSEGASS	SCSETYCGLYPESPEVKAVAS	143					
Db	235	NRMWRKRSFYANNHCIGTDLNRNFASKKWCSEGASS	SCSETYCGLYPESPEVKAVAS	294					
Qy	144	FLRRNINOIKAYISMHSYSHQIVFPYSTRSKSDHEEL	SLVASEAVRAIEKTSKNTRYT	203					

```
Db 295 FLRRNINQIKAYISMHSYSHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 354
Qy 204 HGHGSETLYLAPGGDDMIYDLGIKYSFT 232
Db 355 HGHGSETLYLAPGGDDMIYDLGIKYSFT 383

RESULT 2
US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; PRIOR FILING DATE: 2003-03-04
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Papio hamadryas
US-10-379-836-2

Query Match 87.0%; Score 1164.5; DB 12; Length 423;
Best Local Similarity 80.7%; Pred. No. 1.8e-113;
Matches 217; Conservative 7; Mismatches 8; Indels 37; Gaps 1;

Qy 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPYLYLVKVSKEQTKAKNAI 60
Db 115 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPYLYLVKVSKEQTKAKNAI 174
Qy 61 WIDCGIHAREWISPAFLWFIH----- 83
Db 175 WIDCGIHAREWISPAFLWFIHTEYGIIGYTNLLRHVDYFVMPVNVVDGYDYSWK 234
Qy 84 NRMWRKNSFYANNHCIGTDLNRNFAKWCCEGASSSCSETYCGLYPESPEVKAVAS 143
Db 235 NRMWRKNSFYANNRCIGTDLNRNFAKWCCEGASSSCSETYCGLYPESPEVKAVAN 294
Qy 144 FLRRNINQIKAYISMHSYSHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 295 FLRRNINQIKAYISMHSYSHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 354
Qy 204 HGHGSETLYLAPGGDDMIYDLGIKYSFT 232
Db 355 HGHGSETLYLAPGGDDMIYDLGIKYSFT 383

RESULT 3
US-10-379-836-18
; Sequence 18, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; PRIOR FILING DATE: 2003-03-04
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 18
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Mus musculus
```

```
US-10-379-836-18
Query Match 82.6%; Score 1105.5; DB 12; Length 422;
Best Local Similarity 77.0%; Pred. No. 2.8e-107;
Matches 207; Conservative 10; Mismatches 15; Indels 37; Gaps 1;

Qy 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPYLYLVKVSKEQTKAKNAI 60
Db 114 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPYLYLVKVSKEQTKAKNAI 173
Qy 61 WIDCGIHAREWISPAFLWFIH----- 83
Db 174 WIDCGIHAREWISPAFLWFIHTEYGIIGYTNLLRHVDYFVMPVNVVDGYDYSWK 233
Qy 84 NRMWRKNSFYANNHCIGTDLNRNFAKWCCEGASSSCSETYCGLYPESPEVKAVAS 143
Db 234 NRMWRKNSFYANNRCIGTDLNRNFAKWCCEGASSSCSETYCGLYPESPEVKAVAD 293
Qy 144 FLRRNINQIKAYISMHSYSHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 294 FLRRNINQIKAYISMHSYSHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 353
Qy 204 HGHGSETLYLAPGGDDMIYDLGIKYSFT 232
Db 354 HGHGSETLYLAPGGDDMIYDLGIKYSFT 382

RESULT 4
US-10-379-836-16
; Sequence 16, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; PRIOR FILING DATE: 2003-03-04
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 16
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-379-836-16

Query Match 81.9%; Score 1095.5; DB 12; Length 422;
Best Local Similarity 75.8%; Pred. No. 3.2e-106;
Matches 204; Conservative 11; Mismatches 17; Indels 37; Gaps 1;

Qy 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPYLYLVKVSKEQTKAKNAI 60
Db 114 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSPEKYPYLYLVKVSKEQTKAKNAI 173
Qy 61 WIDCGIHAREWISPAFLWFIH----- 83
Db 174 WIDCGIHAREWISPAFLWFIHTEYGIIGYTNLLRHVDYFVMPVNVVDGYDYSWK 233
Qy 84 NRMWRKNSFYANNHCIGTDLNRNFAKWCCEGASSSCSETYCGLYPESPEVKAVAS 143
Db 234 NRMWRKNSFYANNRCIGTDLNRNFAKWCCEGASSSCSETYCGLYPESPEVKAVAD 293
Qy 144 FLRRNINQIKAYISMHSYSHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
Db 294 FLRRNINQIKAYISMHSYSHIYFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 353
Qy 204 HGHGSETLYLAPGGDDMIYDLGIKYSFT 232
Db 354 HGHGSETLYLAPGGDDMIYDLGIKYSFT 382
```

RESULT 5
US-09-925-302-467
; Sequence 467, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 467
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-467

Query Match 61.0%; Score 816.5; DB 9; Length 211;
Best Local Similarity 90.6%; Pred. No. 2.3e-77;
Matches 154; Conservative 1; Mismatches 6; Indels 9; Gaps 2;

QY 71 WISPAFLWFI-----GHNRMWRKNSFYANNHCIGTDLNRNPFASKHWCCEGASSSS 122
DB 3 WIS-MLCRWLMVMVNSWKQRMWRKNSFYANNHCIGTDLNRNPFASKHWCCEGASSSS 61
QY 123 CSETYGLYPESEPEVKAVASFLRRNQIKAYISMHSYSHQIVFPYSYTRSKSKDHEEL 182
DB 62 CSETYGLYPESEPEVKAVASFLRRNQIKAYISMHSYSHQIVFPYSYTRSKSKDHEEL 121
QY 183 SILVASEAVRAIEKTSKNTYTHGHGSETLYLAPCGGDDWIYDLGIKYSFT 232
DB 122 SILVASEAVRAIEKTSKNTYTHGHGSETLYLAPCGGDDWIYDLGIKYSFT 171

RESULT 6
US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-20

Query Match 42.4%; Score 567; DB 12; Length 402;
Best Local Similarity 41.3%; Pred. No. 8.6e-51;
Matches 111; Conservative 46; Mismatches 72; Indels 40; Gaps 5;

QY 1 ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSKEQTAKNAI 60
DB 96 ATGHSYKKNWKTIEAWTQQVATENPALISRVIGTTEGRAIYLLKV-GRAGQNKPAI 154
QY 61 WIDCGIHAREWISPAFLWFI-----GH----- 83
DB 155 FMDCGFHAREWISPAFLWFI-----GH----- 214
QY 84 NRMWRKNSFYANNHCIGTDLNRNPFASKHWCCEGASSSSCSETYGLYPESEPEVKAVAS 143

DB 215 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCIGISARNPCDETYCGPAASESEKTYKALAD 273
QY 144 FLRRNQIKAYISMHSYSHQIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTY 203
DB 274 FIRNKLSSIKAYLTHTSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 332
QY 204 HGHGSETLYLAPCGGDDWIYDLGIKYSFT 232
DB 333 YGPGATTIYPAAGSGDDWAYDQGIYSFT 361

RESULT 7
US-10-341-434-188
; Sequence 188, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 188
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-188

Query Match 41.3%; Score 552; DB 12; Length 417;
Best Local Similarity 41.4%; Pred. No. 3.4e-49;
Matches 109; Conservative 45; Mismatches 69; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSKEQTAKNAIWDG 65
DB 116 YAKYNNWEKIVAWTEKMDKYPYEMVSRKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCG 174
QY 66 IHAREWISPAFLWFI-----GH-----NRMWR 89
DB 175 IHAREWISPAFLWFI-----GH-----NRMWR 234
QY 89 KNSFYANNHCIGTDLNRNPFASKHWCCEGASSSSCSETYGLYPESEPEVKAVASFLRN 148
DB 235 KNSKQNSKICIGTDLNRNF-NASWNSIPNTNDPCADNTRGSAPESEKTKAVTNFIRSH 293
QY 149 INQIKAYISMHSYSHQIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTYTHGHGS 208
DB 294 LNEIKVYITFHSYSQMLLPYGYTSKLPNHNEDLAKVAKIGTDVL-STRYETRYIYGPIE 352
QY 209 ETLYLAPCGGDDWIYDLGIKYSF 231
DB 353 STIYPISSGSLDWAYDLGIKHTF 375

RESULT 8
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: MPI01-156P1RNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575

PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Fast-Seq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 417
TYPE: PRT
ORGANISM: Homo sapiens
US-10-229-546-2

Query Match 41.3%; Score 552; DB 15; Length 417;
Best Local Similarity 41.4%; Pred. No. 3.4e-49;
Matches 109; Conservative 45; Mismatches 69; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVSKEQTAKNAIWDG 65
DB 116 YAKNWEKIVATERKMDKYPWVRKIKIGSTVEDNPLVLKI-GEKERRKAI FMDCG 174
QY 66 IHAREWISPAFCWLFI-----GH-----NRMR 88
DB 175 IHAREWVSPAFQWVYQATKYGRNKIMTKLLDRMNFYLPVFNVDGYIWSWTKRMR 234
QY 89 KRSFYANNHICIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 148
DB 235 KRSKQNSKICIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPSEKETKAVTNFIRSH 293
QY 149 INQIKAYISMHSYSHIVFPYSTRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
DB 294 LNEIKYIITHFSYSQMLFPYGYTSKLPNHNEDLAKVAKIGTDVL-STRYETRYIYGPIE 352
QY 209 ETLYLAPGGDDWIYDLGIKYSF 231
DB 353 STIYPISSGSLDWAYDLGIKHTF 375

Query Match 41.1%; Score 550.5; DB 9; Length 374;
Best Local Similarity 39.4%; Pred. No. 4.2e-49;
Matches 104; Conservative 45; Mismatches 76; Indels 39; Gaps 4;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVSKEQTAKNAIWDG 65
DB 47 YNIYHPMGEIYEMRREISEKYEVVQTHFLGVTYETHPMYLKI SQSPGNPKKIIMDCG 106
QY 66 IHAREWISPAFCWLFI-----GH-----NRMR 88
DB 107 IHAREWIAFACQWVKEILQNHKONSIRKLLRNLDYVLPVNLIDGYIYTDTDLNR 166
QY 89 KRSFYANNHICIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 148
DB 167 KRSFPHNNGTCFGTDLNRNF-NASWCSIGASRNCODQTFCGTGPVSEPETKAVASFIESK 225
QY 149 INQIKAYISMHSYSHIVFPYSTRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
DB 226 KDDILCLFTHSYGQLILTPYGYTKNSNHPHMQVQKAAANAL-KAKYGTNYRYGSSA 284
QY 209 ETLYLAPGGDDWIYDLGIKYSF 232
DB 285 DILYASSGSSRDWARDIGIPFSIT 308

RESULT 10
US-09-888-615-61
Sequence 61, Application US/09888615
Patent No. US20020064856A1
GENERAL INFORMATION:
APPLICANT: PLOWMAN, GREGORY
APPLICANT: WHITE, DAVID
APPLICANT: CAENEPEEL, SEAN
APPLICANT: CHARYDCZAK, GLEN
APPLICANT: MANNING, GERARD
APPLICANT: SUDARSANAM, SUCHA
TITLE OF INVENTION: NOVEL PROTEASES
FILE REFERENCE: 038602/1214
CURRENT APPLICATION NUMBER: US/09/888,615
CURRENT FILING DATE: 2001-06-26
PRIOR APPLICATION NUMBER: 60/214,047
PRIOR FILING DATE: 2000-06-26
NUMBER OF SEQ ID NOS: 150
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 61
LENGTH: 374
TYPE: PRT
ORGANISM: Homo sapiens
US-09-888-615-61

PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Fast-Seq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 417
TYPE: PRT
ORGANISM: Homo sapiens
US-10-229-546-2

Query Match 41.3%; Score 552; DB 15; Length 417;
Best Local Similarity 41.4%; Pred. No. 3.4e-49;
Matches 109; Conservative 45; Mismatches 69; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVSKEQTAKNAIWDG 65
DB 116 YAKNWEKIVATERKMDKYPWVRKIKIGSTVEDNPLVLKI-GEKERRKAI FMDCG 174
QY 66 IHAREWISPAFCWLFI-----GH-----NRMR 88
DB 175 IHAREWVSPAFQWVYQATKYGRNKIMTKLLDRMNFYLPVFNVDGYIWSWTKRMR 234
QY 89 KRSFYANNHICIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 148
DB 235 KRSKQNSKICIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPSEKETKAVTNFIRSH 293
QY 149 INQIKAYISMHSYSHIVFPYSTRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
DB 294 LNEIKYIITHFSYSQMLFPYGYTSKLPNHNEDLAKVAKIGTDVL-STRYETRYIYGPIE 352
QY 209 ETLYLAPGGDDWIYDLGIKYSF 231
DB 353 STIYPISSGSLDWAYDLGIKHTF 375

Query Match 41.1%; Score 550.5; DB 9; Length 374;
Best Local Similarity 39.4%; Pred. No. 4.2e-49;
Matches 104; Conservative 45; Mismatches 76; Indels 39; Gaps 4;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPYLVKVSKEQTAKNAIWDG 65
DB 47 YNIYHPMGEIYEMRREISEKYEVVQTHFLGVTYETHPMYLKI SQSPGNPKKIIMDCG 106
QY 66 IHAREWISPAFCWLFI-----GH-----NRMR 88
DB 107 IHAREWIAFACQWVKEILQNHKONSIRKLLRNLDYVLPVNLIDGYIYTDTDLNR 166
QY 89 KRSFYANNHICIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 148
DB 167 KRSFPHNNGTCFGTDLNRNF-NASWCSIGASRNCODQTFCGTGPVSEPETKAVASFIESK 225
QY 149 INQIKAYISMHSYSHIVFPYSTRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
DB 226 KDDILCLFTHSYGQLILTPYGYTKNSNHPHMQVQKAAANAL-KAKYGTNYRYGSSA 284
QY 209 ETLYLAPGGDDWIYDLGIKYSF 232
DB 285 DILYASSGSSRDWARDIGIPFSIT 308

RESULT 11
US-10-176-306-74
Sequence 74, Application US/10176306
Publication No. US20030130485A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel A.
APPLICANT: Curtis, Rory A. J.
APPLICANT: Glucksman, Maria Alexandra
APPLICANT: Bandaru, Rajasekhar
APPLICANT: Kapeller-Liebermann, Rosana
TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
FILE REFERENCE: 10448-195001
CURRENT APPLICATION NUMBER: US/10/176,306
CURRENT FILING DATE: 2002-06-20

```

; PRIOR APPLICATION NUMBER: 10/001,137
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: PCT/US01/45291
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/248,362
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,331
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,365
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/250,077
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,327
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,176
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 10/023,617
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/49416
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,249
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,405
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 10/083,248
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/US01/46717
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242,324
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,518
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-306-74

Query Match          41.1%; Score 550.5; DB 16; Length 444;
Best Local Similarity 39.3%; Pred. No. 5.3e-49;
Matches 104; Conservative 45; Mismatches 76; Indels 39; Gaps 4;

QY      6 YEOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEYPLVYLVKVSQKEQAKNAIWDG 65
Db      177 YNIYHPGELIYEWREISEKYEVTQHFGLGVTYTHPIYLYKISQPSGNPKLIWWDG 236

QY      66 IHAREWISPAFCIWF-
Db      237 IHAREWIAFAFCQFWKYEILQNHKNSRIRKLRNLDFYVLPVNLIDGYIYTTDLRWR 296

QY      89 KRSFYANNHCIGTDLNRNPFASKHWCCEGSSSCSETYCGLYPESEPEVKAVASFLRN 148
Db      297 KRSRPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTCGTPGVSEPETKAVASFTESK 355

QY      149 INQIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db      356 KDDILCFLTWHHSGQLTLTPYGYTKNKSNNHPMIQVQKAAANAL-KAKYGTNYRVGSSA 414

QY      209 ETLYLAPGGDDWIYDLGIKYSFT 232
Db      415 DILYASSGSRDRNARDIGIFPSVT 438

RESULT 12
US-10-274-639-12
; Sequence 12, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
```

```

; APPLICANT: DELEGEANE, Angelo M.; GANDHI, Ameena R.
; APPLICANT: HAPALIA, April J.A.; LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
; APPLICANT: NGUYEN, Damiel B.; LEE, Ernestine A.
; APPLICANT: KHAN, Farrah A.; YUE, Henry
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha
; APPLICANT: DING, Li; KEARNEY, Liam
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.
; APPLICANT: LAL, Preeti G.; LEE, Sally
; APPLICANT: TODD, Stephen; LO, Terence P.
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,717
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/225,988
; PRIOR FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 60/227,568
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 12
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

Query Match          41.0%; Score 549; DB 12; Length 437;
Best Local Similarity 40.3%; Pred. No. 7.5e-49;
Matches 106; Conservative 40; Mismatches 77; Indels 40; Gaps 4;

QY      6 YEOYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEYPLVYLVKVSQKEQAKNAIWDG 65
Db      136 YEVYHSLSEIQNWHLNTHSLGIHMFSGRSGVGRSLFILKL-GRRSLKRAVWIDCG 194

QY      66 IHAREWISPAFCIWF-
Db      195 IHAREWIGPAFCQFWKYEALLTYKSDPAMKMLNHLFYIMPVENVVDGYHFSWTDNDFWR 254

QY      89 KRSFYANNHCIGTDLNRNPFASKHWCCEGSSSCSETYCGLYPESEPEVKAVASFLRN 148
Db      255 KTRNSRNRFRRCRGVDANRNKVK-WCEGASMHPCDDTCGPPFSEPEVKAVANFURKH 313

QY      149 INQIKAYISMHSYSHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db      314 RKHIRAYLSFHAYAOMLLYPYSYKYATIPNFRCVESNAYKAVNALQSV-YGVTRYVGPAS 372

QY      209 ETLYLAPGGDDWIYDLGIKYSF 231
Db      373 TTLVSSGSSMDWAYKNGIPYAF 395

RESULT 13
US-10-200-344-10
```



```
; Sequence 10, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
; FILE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 437
; TYPE: PRT
; ORGANISM: homo sapiens
; US-10-200-344-10

Query Match 41.0%; Score 549; DB 14; Length 437;
Best Local Similarity 40.3%; Pred. No. 7.5e-49;
Matches 106; Conservative 40; Mismatches 77; Indels 40; Gaps 4;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLVLYKVSQKEQTAKNAIWDG 65
DB 136 YEVVHSLLEEIQNMWHLKNTHSGLIHMFSGISYEGRSFLIKL-GRRSLKRAVWDG 194
QY 66 IHAREWISPAFLWFI-----GHNRMWR 88
DB 195 IHAREWISPAFLWFI-----GHNRMWR 254
QY 89 KNSFYANNHCIGTDLNRFASKHWCESGSSSCSETYCGLYPESPEVKAVASFLRRN 148
DB 255 KTRSRNRFRCRGVDANRNWKV-WCEGASMPDCDTCYGPPESEPEVKAVANFLRKH 313
QY 149 INQIKAYISMHSYQHSQHYVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHS 208
DB 314 RKHRAVLSFHAYAQMLLYPSYKYATIPNRCVESAAYKAVNALQSV-YGVRYRYGPAS 372
QY 209 ETLYLAPGGDDWIYDLGIKYSPT 231
DB 373 TTYLVSSGSSMDWAYKNGIPYAF 395

RESULT 14
US-09-910-059-113
; Sequence 113, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; FILE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113

Query Match 41.0%; Score 549; DB 10; Length 716;
Best Local Similarity 40.1%; Pred. No. 1.5e-48;
Matches 108; Conservative 47; Mismatches 74; Indels 40; Gaps 5;

QY 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLVLYKVSQKEQTAKNAI 60
DB 111 ATGHSYKYNKWTIEAWTQQVATENPALISRSVIGTFEGRAIYLLKV-GKAGQNKPAI 169
QY 61 WIDCGIHAREWISPAFLWFI-----GH----- 83
; SEQ ID NO 113
```

```
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: humanised Fd mutant HCPB sequence
; US-09-910-059-113

Query Match 41.0%; Score 549; DB 10; Length 613;
Best Local Similarity 40.1%; Pred. No. 1.2e-48;
Matches 108; Conservative 47; Mismatches 74; Indels 40; Gaps 5;

QY 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLVLYKVSQKEQTAKNAI 60
DB 307 ATGHSYKYNKWTIEAWTQQVATENPALISRSVIGTFEGRAIYLLKV-GKAGQNKPAI 365
QY 61 WIDCGIHAREWISPAFLWFI-----GH----- 83
DB 366 FMDCGFHAREWISPAFLWFI-----GH----- 425
QY 84 NRMWRKNSFYANNHCIGTDLNRFASKHWCESGSSSCSETYCGLYPESPEVKAVAS 143
DB 426 SRFWRKTRSTHTGSSCIGTDPNRF-DAGWCEIGASRNPCDETYCGPAASEKETKALAD 484
QY 144 FLRNINQIKAYISMHSYQHSQHYVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 485 FINKLSSIKAYLTIHSYQSMIYPIYSYAYKLGNNALNALAKATVKEL-ASLHGTYT 543
QY 204 HGHSETLYLAPGGDDWIYDLGIKYSPT 232
DB 544 YGPGATTIYPSAGTSKDMWAYDQGIYSFT 572

RESULT 15
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
; APPLICANT: Copley, Clive G
; APPLICANT: Edge, Michael Derek
; APPLICANT: Emery, Stephen Charles
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
; FILE OF INVENTION: Their Therapeutic use in an Adept System
; FILE REFERENCE: 1991-209
; CURRENT APPLICATION NUMBER: US/09/910,059
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 09/171,945
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; PRIOR APPLICATION NUMBER: GB 9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125

Query Match 41.0%; Score 549; DB 10; Length 716;
Best Local Similarity 40.1%; Pred. No. 1.5e-48;
Matches 108; Conservative 47; Mismatches 74; Indels 40; Gaps 5;

QY 1 ASASYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLVLYKVSQKEQTAKNAI 60
DB 111 ATGHSYKYNKWTIEAWTQQVATENPALISRSVIGTFEGRAIYLLKV-GKAGQNKPAI 169
QY 61 WIDCGIHAREWISPAFLWFI-----GH----- 83
; SEQ ID NO 125
```

Db	170	FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDPYLVPLVINIDGYIYTWTK	229
Qy	84	NRMRKNSFYANNHCICTDLNRNPFASKHWCBEGASSSSCSSTYCGLYPESEPEVKAVAS	143
Db	230	SREWRKTRSTHTGSSCICGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD	288
Qy	144	PLRNNIQIKAYISMHSYSHIYVPYSTRSKSDHEELSLVASBAVRAIEKTSKNTRYT	203
Db	289	FIRKLSIKAYLTHYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTYKT	347
Qy	204	HGHGSETLYLAPGGGDDWIYDLGIKYSFT	232
Db	348	YPGGATTIYPSAGTSKOWAYDQGRYSFT	376

Search completed: January 1, 2004, 19:50:59
Job time : 36.1795 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:25:56 ; Search time 16.9478 Seconds
(without alignments)
1395.903 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQVHSLNEIYSWIE.....IKYFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1239.5	92.6	423	2 A41204	carboxypeptidase B
2	601	44.9	415	2 A32129	carboxypeptidase B
3	586	43.8	306	1 CP80B	carboxypeptidase B
4	561	41.9	417	1 A34487	carboxypeptidase A
5	552	41.3	309	2 A38395	carboxypeptidase A
6	552	41.3	417	2 A43929	carboxypeptidase A
7	547.5	40.9	416	1 A42332	carboxypeptidase A
8	502.5	37.6	419	1 CP80A	carboxypeptidase A
9	487.5	36.4	419	1 CP81A	carboxypeptidase A
10	483	36.1	417	2 A32128	carboxypeptidase A
11	465.5	34.8	419	1 S29127	carboxypeptidase A
12	456	34.1	417	2 A56171	carboxypeptidase A
13	403	30.1	323	2 T20507	hypothetical prote
14	396.5	29.6	528	2 T33527	hypothetical prote
15	375.5	28.1	303	1 CPCVB	carboxypeptidase B
16	366.5	27.4	455	2 T33572	hypothetical prote
17	363	27.1	527	2 T27572	hypothetical prote
18	344.5	25.7	667	2 T33526	hypothetical prote
19	342	25.6	666	2 T24170	hypothetical prote
20	286.5	21.4	424	2 S17571	carboxypeptidase T
21	286.5	21.4	430	2 S48976	ECM14 protein - ye
22	282.5	21.1	451	2 S20723	carboxypeptidase (
23	250.5	18.7	373	2 T26030	hypothetical prote
24	246	18.4	497	2 T40260	carboxypeptidase p
25	177.5	13.3	999	2 T36021	probable zinc-bind
26	139.5	10.4	558	2 S75104	hypothetical prote
27	131.5	9.8	1446	2 T30916	carboxypeptidase D
28	109.5	8.2	985	2 T29910	hypothetical prote
29	101	7.5	1216	2 H84629	hypothetical prote

30	99	7.4	444	2 S51199	tryptophan 5-monoo
31	99	7.4	491	2 B96739	hypothetical prote
32	97	7.2	444	1 S10489	tryptophan 5-monoo
33	96.5	7.2	376	1 E69957	gamma-D-glutamyl-L
34	96	7.2	1389	2 I50090	carboxypeptidase g
35	95	7.1	707	2 T40070	origin recognition
36	94	7.0	477	2 S16383	carboxypeptidase E
37	93.5	7.0	355	2 C83850	gamma-D-glutamyl-L
38	92	6.9	985	2 D82776	pyruvate dehydroge
39	91	6.8	262	2 A64882	probable carboxype
40	91	6.8	262	2 A99867	probable carboxype
41	91	6.8	262	2 H85751	probable carboxype
42	90.5	6.8	448	2 A24327	carboxypeptidase E
43	90.5	6.8	476	2 S12461	carboxypeptidase E
44	90.5	6.8	476	2 A40469	carboxypeptidase E
45	89.5	6.7	1119	2 T13284	carboxypeptidase (

ALIGNMENTS

RESULT 1

A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N:Alternate names: plasma carboxypeptidase B
C:Species: Homo sapiens (man)
C:Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C:Accession: A41204
R:Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A:Title: Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase
A:Reference number: A41204; MOID:92042093; PMID:1939207
A:Accession: A41204
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-423 <EAT>
A:Cross-references: GB:M75106; NID:gl89686; PID:gl89687
C:Genetics:
A:Gene: GDB:CPB2
A:Cross-references: GDB:129546; OMIM:212070
A:Map position: 13q14.11-13q14.11
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-114/Domain: activation peptide #status predicted <ACP>
F:115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Query Match 92.6%; Score 1239.5; DB 2; Length 423;
Best Local Similarity 86.2%; Pred. No. 7.5e-104;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY	1	ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYVLYLVKSGKEQTAKNAI	60
DB	115	ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHGSSPEKPYVLYLVKSGKEQTAKNAI	174
QY	61	WIDCGIHAREWISPAFLWFIGH-----	83
DB	175	WIDCGIHAREWISPAFLWFIGHITQFYGIQGYTNLLRLVDFYVMPVNVVDGYDYSWK	234
QY	84	NRMWRKNRNFYANNHCIGTDLNRNFKSKWCEGASSSSCSETYCGLYPSEPEVKAVAS	143
DB	235	NRMWRKNRNFYANNHCIGTDLNRNFKSKWCEGASSSSCSETYCGLYPSEPEVKAVAS	294
QY	144	FLRRNINQIKAVISMHSYSQHLVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT	203
DB	295	FLRRNINQIKAVISMHSYSQHLVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT	354
QY	204	HGHGSETLYLAPGGDDWIYDLGIKYSFT	232
DB	355	HGHGSETLYLAPGGDDWIYDLGIKYSFT	383

RESULT 2

J. Mol. Biol. 103, 175-190, 1976

A:Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A:Reference number: A92051; MUID:76265065; PMID:957425
A:Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R:Plummer Jr., T.H.

J. Biol. Chem. 244, 5246-5253, 1969
A:Title: Isolation and sequence of peptides at the active center of bovine carboxypeptidase B
A:Reference number: A92051; MUID:70007159; PMID:5344132
A:Contents: annotation; active site
R:Kimmel, M.T.; Plummer Jr., T.H.

J. Biol. Chem. 247, 7864-7869, 1972
A:Title: Identification of a glutamic acid at the active center of bovine carboxypeptidase B
A:Reference number: A92119; MUID:73061487; PMID:4565668
A:Contents: annotation; active site
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:63-76,135-159,149-163/Disulfide bonds: #status experimental
F:66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F:246,268/Active site: Tyr, Glu #status experimental

Query Match 43.8%; Score 586; DB 1; Length 306;
Best Local Similarity 43.9%; Pred. No. 3.8e-45;
Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

Qy 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPLYLVKVGKEQTAKNAIWDG 65
Db 6 YEKYNNWETIEAWTEQVASENPDLISRAIGTTPLGNTIYLLKV-GKPGSNKPAVFMDCG 64

Qy 66 IHAREWISPAFLWFI-----GH-----NMWR 88
Db 65 FHAREWISPAFCQWFREAVRTYGREIHMTEPLDKLDFVLPVAVNIDGVIYTTWTTNRMR 124

Qy 89 KNRSFYANNHCIGTDLNRNFASKWCCEGAGSSSCSYTCGLYPESEPKVAVASFLRN 148
Db 125 KTRSTRAGSSCTGTDLNKNF-DAGWCSTGASNPCEYTCGSAASEKSKAVADFIRNH 183

Qy 149 INQIKAYISMHSYQHIYFPYSYTRSKSKDHLESLVASEAVRAIEKTSKNTRYTHGHGS 208
Db 184 LSSIKAVLTIHYSQMMLYPYSDYKLPKNVVELNLAKGAVKKL-ASLHGITYSYGPGA 242

Qy 209 ETLVAPGGGDDWIYDLGIKYSFT 232
Db 243 TTIYPASGGSDWAYDQGIKYSFT 266

RESULT 4
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A34487
R:Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A:Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member
A:Reference number: A34487; MUID:90062123; PMID:2584208
A:Accession: A34487
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-417 <REX>
A:Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:G309135
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase
F:356,378/Active site: Tyr, Glu #status predicted

Query Match 41.9%; Score 561; DB 1; Length 417;
Best Local Similarity 43.7%; Pred. No. 1e-42;
Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;

Qy 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHGSSPEKYPLYLVKVGKEQTAKNAIWDG 65
Db 116 YAKYNDWDKIYSWTETKMLEKHPWMSRIKIGSTDNPDLVILKI-GKDGKRAIFMDCG 174

Qy 66 IHAREWISPAFLWFI-----GHN-----RMWR 88

Db 175 IHAREWISPAFCWFYQATKSGKKNIMTKLLDRNMFYVLPVFNVDGYIWSWTKRMWR 234
QY 89 KRSFYANNHCIGTDLNRNPFASKHWCCEGAGSSSCSETCYGLYPESEPEVKAVASFLRRN 148
Db 235 KNSRQNSQTCIGTDLNRNF-DVSDSSPNTNPKCLNVYRGPAPESEKETKAVTNFIRSH 293
QY 149 INQIKAYISMHSYSHIVPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db 294 LNSIKAYITFHSYSQMLLPYGYTKLPNQHDLKLVARIATDAL-STRYETRYIYGPIA 352
QY 209 ETLYLAPGGDDWIYDLGIKYSF 231
Db 353 STIYKTSGLSDWAYDLGIKHTF 375
RESULT 5
A38395
mest cell carboxypeptidase (EC 3.4.17.1) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C:Accession: A38395; A33118
R: Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A: Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme activation
A: Reference number: A38395; PMID:9105153; PMID:1988052
A: Accession: A38395
A: Status: preliminary
A: Molecule type: protein
A: Residues: 1-309 <COL>
R: Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A: Reference number: A33118
A: Accession: A33118
A: Status: preliminary
A: Molecule type: protein
A: Residues: 2-309 <CO2>
C: Superfamily: carboxypeptidase
C: Keywords: hydrolase
F: 248.270/Active site: Tyr, Glu #status predicted

Query Match 41.3%; Score 552; DB 2; Length 309;
Best Local Similarity 42.6%; Pred. No. 4.5e-42;
Matches 112; Conservative 39; Mismatches 72; Indels 40; Gaps 5;
QY 6 YEOYHSLNIYSWIEPIERHPDMLTKIHIGSFEKYPLVYLVKSGEQTAKNAIWDG 65
Db 8 YAKYNDWKIVSTKQWEKHPMVSRIKIGSTVEDNPLYLKI-GRKDGKRAIFMDCG 66
QY 66 IHAREWISPAFCWLFI-----GHN-----RMWR 88
Db 67 IHAREWVSPAFQCFWYQAQKSGKKNIMTKLLDRNMFYVLPVFNVDGYIWSWTKRMWR 126
QY 89 KRSFYANNHCIGTDLNRNPFASKHWCCEGAGSSSCSETCYGLYPESEPEVKAVASFLRRN 148
Db 127 KRSRQNSQTCIGTDLNRNF-DVSDSSPNTNPKCLNVYRGPAPESEKETKAVTNFIRSH 185
QY 149 INQIKAYISMHSYSHIVPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db 186 LNSIKAYITFHSYSQMLLPYGYTKLPNQHDLKLVARIATDAL-SSRYETRYIYGPIA 244
QY 209 ETLYLAPGGDDWIYDLGIKYSF 231
Db 245 STIYKTSGLSDWAYDLGIKHTF 267
RESULT 6
A43929
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N: Alternate names: mast cell carboxypeptidase A3
C: Species: Homo sapiens (man)
C: Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C: Accession: A43929; A39246; A45759

Db 175 IHAREWISPAFCWFYQATKSGKKNIMTKLLDRNMFYVLPVFNVDGYIWSWTKRMWR 234
QY 89 KRSFYANNHCIGTDLNRNPFASKHWCCEGAGSSSCSETCYGLYPESEPEVKAVASFLRRN 148
Db 235 KNSRQNSQTCIGTDLNRNF-DVSDSSPNTNPKCLNVYRGPAPESEKETKAVTNFIRSH 293
QY 149 INQIKAYISMHSYSHIVPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db 294 LNSIKAYITFHSYSQMLLPYGYTKLPNQHDLKLVARIATDAL-STRYETRYIYGPIA 352
QY 209 ETLYLAPGGDDWIYDLGIKYSF 231
Db 353 STIYKTSGLSDWAYDLGIKHTF 375
RESULT 7
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N: Alternate names: carboxypeptidase B isozyme 1
C: Species: Homo sapiens (man)
C: Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C: Accession: A42332; S02824; S02813; S02812
R: Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Shami, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992

R: Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A: Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A
A: Reference number: A43929; PMID:92105393; PMID:1729276
A: Accession: A43929
A: Molecule type: DNA
A: Residues: 1-417 <REY>
A: Cross-references: GB:M73716
A: Experimental source: mast cell
A: Note: The authors translated the codon CCG for residue 231 as Thr
R: Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugraber, D.J.; Austen, K.F.; Serafin, V.
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A: Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison
A: Reference number: A39246; PMID:9083291; PMID:2594780
A: Accession: A39246
A: Molecule type: mRNA
A: Residues: 1-417 <RE2>
A: Cross-references: GB:M27717; NID:gl79933; PIDN:AAA35652.1; PID:gl79934
R: Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.U.
J. Clin. Invest. 83, 1630-1636, 1989
A: Title: Human mast cell carboxypeptidase. Purification and characterization.
A: Reference number: A45759; PMID:89214692; PMID:2708524
A: Accession: A45759
A: Molecule type: protein
A: Residues: 110-137 <GOL>
C: Genetics:
A: Gene: GDB:CPA3
A: Cross-references: GDB:125231; OMIM:114851
A: Map position: 3q21.3-3q25
A: Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C: Superfamily: carboxypeptidase
C: Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc.
F: 115/Domain: signal sequence #status predicted <SIG>
F: 116-109/Domain: activation peptide #status predicted <ACT>
F: 110-417/Product: carboxypeptidase A, mast cell #status predicted <MAT>
F: 176.179.304/Binding site: zinc (His, Glu, His) #status predicted
F: 245-268/Disulfide bonds: #status predicted
F: 356.378/Active site: Tyr, Glu #status predicted

Query Match 41.3%; Score 552; DB 2; Length 417;
Best Local Similarity 41.4%; Pred. No. 6.5e-42;
Matches 109; Conservative 45; Mismatches 69; Indels 40; Gaps 5;
QY 6 YEOYHSLNIYSWIEPIERHPDMLTKIHIGSFEKYPLVYLVKSGEQTAKNAIWDG 65
Db 116 YAKYNNWEKIVAWTKMDKYPMSRIKIGSTVEDNPLYLKI-GEKNERRKAIFMDCG 174
QY 66 IHAREWISPAFCWLFI-----GH-----RMWR 88
Db 175 IHAREWVSPAFQCFWYQAQKSGKKNIMTKLLDRNMFYVLPVFNVDGYIWSWTKRMWR 234
QY 89 KRSFYANNHCIGTDLNRNPFASKHWCCEGAGSSSCSETCYGLYPESEPEVKAVASFLRRN 148
Db 235 KRSRQNSQTCIGTDLNRNF-NASWNSIPTNDPCADNTRGSAPESEKETKAVTNFIRSH 293
QY 149 INQIKAYISMHSYSHIVPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db 294 LNEIKAYITFHSYSQMLLPYGYTKLPNQHDLKLVARIATDAL-STRYETRYIYGPIE 352
QY 209 ETLYLAPGGDDWIYDLGIKYSF 231
Db 353 STIYKTSGLSDWAYDLGIKHTF 375
RESULT 7
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N: Alternate names: carboxypeptidase B isozyme 1
C: Species: Homo sapiens (man)
C: Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C: Accession: A42332; S02824; S02813; S02812
R: Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Shami, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992

F:358,380/Active site: Tyr, Glu #status experimental

Query Match 37.6%; Score 502.5; DB 1; Length 419;
Best Local Similarity 40.2%; Pred. No. 1.9e-37;
Matches 106; Conservative 42; Mismatches 75; Indels 41; Gaps 5;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEXKPYLVLVKVSKEQTAKNAIWDG 65
DB 119 YATHILDEIYDFMLDLVAEHPOLVSKIQGRSIEGRPIYVLKFS-TGGSNRPAIWDLG 177
QY 66 IHAREWISPAFCLEWFI-----GHNRMWR 88
DB 178 IHSREMITQATGVWFAKKFTEDYQDPSTAILDSMDIFLEIVTNPDPFAFTHSQNRLMR 237
QY 89 KNRSPYANNHCITGTDLNRNPFASKHWECEGASSSCSETTCGLYPESEPEVKAVASFLRN 148
DB 238 KTRSTVTSLLCVGVDPNRNW-DAGFGKAGASSPCSETYHGKVSANSEVEVKISVDFVKDH 296
QY 149 INQIKAYISMHSYQSHIVPYPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHGS 208
DB 297 GN-FKAPLSIHSYQLLYPYGYTSEPAPDQAELOLAKSANTAL-TSLHGTEFKY 350
QY 209 ETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 TTIYQASGSDSYNOGIKYSFT 378

RESULT 9

CPRTA

carboxypeptidase A (EC 3.4.17.1) precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999

C:Accession: A00911; B32129

R:Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standring, D.N.; Pictet, R.I. Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982

A:Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of A:Reference number: A00911; MUID:82105986; PMID:6275388

A:Accession: A00911

A:Molecule type: mRNA

A:Residues: 1-419 <GUI>

A:Cross-references: GB:J00713; NID:G203364; PIDN:AAA0893.1; PID:G203365

R:Clauer, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.

J. Biol. Chem. 263, 17837-17845, 1988

A:Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparison of A:Reference number: A92693; MUID:89034324; PMID:3182872

A:Accession: B32129

A:Molecule type: DNA

A:Residues: 1-260, 'P', 262, 'M', 264-346, 'K', 348-419 <CLA>

C:Genetics:

A:Gene: CPA1

A:Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1

A>Note: the authors translated the codon GTC for residue 196 as Ile

C:Superfamily: carboxypeptidase

C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-110/Domain: activation peptide #status predicted <PRO>

F:111-419/Product: carboxypeptidase A #status predicted <MAT>

F:179,182,306/Binding site: zinc (His, Glu, His) #status predicted

F:248-271/Diulfide bonds: #status predicted

F:358,380/Active site: Tyr, Glu #status predicted

Query Match 36.4%; Score 487.5; DB 1; Length 419;

Best Local Similarity 39.2%; Pred. No. 4.2e-36;

Matches 105; Conservative 40; Mismatches 74; Indels 49; Gaps 6;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEXKPYLVLVKVSKEQTAKNAIWDG 65

DB 119 YATHILDEIYDFMLDLVAEHPOLVSKIQIGNTFGRPIHVLKFS-TGGSNRPAIWDTG 177

QY 66 IHAREWISPAFCLEWFI-----GHNRMWR 88

DB 178 IHSREMITQATGVWFAKKFTEDYQDPSTAILDSMDIFLEIVTNPDPFAFTHTKTRNMR 237

QY 89 KNRSPYANNHCITGTDLNRNPFASKHWECEGASSSCSETTCGLYPESEPEVKAVASFLRN 144
DB 238 KTRSTVTSLLCVGVDPNRNW-DAGFGKAGASSPCSETYHGKVSANSEVEVKISVDFVKDH 292
QY 145 LRNNINQIKAYISMHSYQSHIVPYPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTH 204
DB 293 VTSHGN-IKAFISIHSYQLLYPYGYTSEPAPDQAELOLAKSANTAL-TSLHGTEFKY 350
QY 205 GHGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 351 GSIIDTIYQASGSDSYNOGIKYSFT 378

RESULT 10

A32128

carboxypeptidase A2 (EC 3.4.17.15) precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999

C:Accession: A32128

R:Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rutter, J. Biol. Chem. 263, 17828-17836, 1988

A:Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and expression of A:Reference number: A32128; MUID:89034323; PMID:3182871

A:Accession: A32128

A:Molecule type: mRNA

A:Residues: 1-417 <GAR>

A:Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:G341029; PIDN:AAA0956.1; PID:G341029

C:Superfamily: carboxypeptidase

C:Keywords: hydrolase; metallo-carboxypeptidase

F:356,378/Active site: Tyr, Glu #status predicted

Query Match 36.1%; Score 483; DB 2; Length 417;

Best Local Similarity 38.2%; Pred. No. 1.1e-35;

Matches 109; Conservative 47; Mismatches 75; Indels 54; Gaps 9;

QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEXKPYLVLVKVS-GKEQTAKNAIWDG 64

DB 118 FEAYHTLEIYQEMDLVAENPGLVSKNLGSSFENRPMNVLPKFTGGD---KPAIWLDA 174

QY 65 GHAREWISPAFCLEWFI-----FI-----GH-----NRMW 87

DB 175 GHAREWISPAFCLEWFI-----FI-----GH-----NRMW 234

QY 88 KNRSPYANNHCITGTDLNRNPFASKHWECEGASSSCSETTCGLYPESEPEVKAVASFLRN 147

DB 235 RKTRSKRSGCGVDPNRNW-DANFGPGASSPCSDSYHGKPNSEVEVKISVDFIKS 293

QY 148 NINQIKAYISMHSYQSHIVPYPYSTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHG 207

DB 294 H-GKVKAFITLHSYQLLYPYGYTSEPAPDQAELOLAKSANTAL-KRLHGTSTYKVGPI 351

QY 208 SETLYLAPGGDDWIYDLGIKYSFTSN-----PPVEKLLP 242

DB 352 CSVIYQASGSDSYNOGIKYSFTSN-----PPVEKLLP 396

RESULT 11

S29127

carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human

N:Alternate names: pancreatic carboxypeptidase A1

C:Species: Homo sapiens (man)

C>Date: 25-Feb-1994 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999

C:Accession: S29127; A34205; S08253; S02810; S71394; S02811

R:Cataaus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver, Biochem. J. 287, 299-303, 1992

A:Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1. A:Reference number: S29127; MUID:93038569; PMID:1417781

A:Accession: S29127

A:Molecule type: mRNA

A:Residues: 1-419 <CAT>

A:Cross-references: EMBL:X67318; NID:G953329; PIDN:CAA47732.1; PID:G953330

R:Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.

Am. J. Hum. Genet. 46, 795-800, 1990

A;Title: Human carboxypeptidase A identifies a BgIII RFLP and maps to 7q31-qter.
A;Reference number: A34205; MUID:90196012; PMID:1969228
A;Accession: A34205
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 330-396 <STS>
A;Note: the authors translated the codon CTG for residue 391 as Val
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxypeptidase A
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08253
A;Molecule type: protein
A;Residues: 17-43; 'XXX', 114-135 <MOU>
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase A
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A;Molecule type: protein
A;Residues: 17-42 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton, Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic procarboxypeptidase A1 and A2
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71394
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-419 <LAE>
C;Genetics:
A;Gene: GDB:CPA1; CPA
A;Cross-references: GDB:120597; OMIM:114850
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-protein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

Query Match 34.8%; Score 465.5; DB 1; Length 419;
Best Local Similarity 37.9%; Pred. No. 4.1e-34;
Matches 100; Conservative 44; Mismatches 79; Indels 41; Gaps 5;

QY 6 YEQHSLSNEYSWIEFITERHPDMLTKIHIGSSFEKYPYLVLKVSQKQAKNAIWDCG 65
Db 119 YATVHTLEEIYDFLDLLVAENPHLVSKIQIGNTVGRPIYVLKFS-TGGSKRPAIWDTG 177
QY 66 IHAREWISPAFCWFI-----GHNRMWR 88
Db 178 IHSREWVTOAGVWFVAKKITQDYQDAAFTAILDLTDLIFLEIVTNPDPGFATHTSNRMWR 237
QY 89 KNSFYANNHCIGTDLNRNFASKHWCCEGASSSCSCTCYGLYPESEPEVKAVASFLRN 148
Db 238 KTSHTAGSLCIGVDPNRNWDAEGL-SGASSPCSTYHGKFSANSEVEKSVDFVKDH 296
QY 149 INQIKAYISMHSYQSHIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
Db 297 GN-1KAFISHSYQLLMFPYGYKCTKLDLDFELSEVAQAQSL-RSLHGTYKYGKVICSV 354
QY 209 ETYLAPGGDDWIYDLGKYSFT 232
Db 355 KAIYQAGSGTIDWTYSQGIKYSFT 378

RESULT 12

AS6171
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999

C;Accession: A56171; S02809; S71395
R;Cataus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M. J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. CDNA
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:G790226; PIDN:AAA74425.1; PID:G790227
A;Note: authors translated the codon AGA for residue 339 as Ser, and ACC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase A
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: 'S', 18-36; 'N', 38-43 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton, Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic procarboxypeptidase A1 and A2
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-301; 'T', 303-338; 'SR', 341-417 <LAE>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Query Match 34.1%; Score 456; DB 2; Length 417;
Best Local Similarity 36.5%; Pred. No. 2.9e-33;
Matches 103; Conservative 45; Mismatches 80; Indels 54; Gaps 7;

QY 9 YHSLSNEYSWIEFITERHPDMLTKIHIGSSFEKYPYLVLKVS-GKEQTAKNAIWDCGIH 67
Db 121 YHLEISQEMDLVAEHPCLVSKVNGISFENRPMNLKFSITGGD---KPAIWDAGI 177
QY 68 AREWISPAFCWFI-----HNRMWRKN 90
Db 178 AREWVTOATALTANKIVSDYKDPISITSLDALDIFLLPVTNPDPGVFSQTKRMWRKT 237
QY 91 RSYFANNHCIGTDLNRNFASKHWCCEGASSSCSCTCYGLYPESEPEVKAVASFLRNIN 150
Db 238 RSKVSGSLCVGVDPNRNW-DAGFGPGCASNPCSDSYHGSPANSSEVEKSVDFIKSH-G 295
QY 151 QIKAYISMHSYQSHIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGHSET 210
Db 296 KVAFIILHSYQLLMFPYGYKCTKLDLDFELSEVAQAQSL-RSLHGTYKYGKVICSV 354
QY 211 LYLAPGGDDWIYDLGKYSFTSN-----PPVEKLLP 242
Db 355 IYQAGSGSIDWSYDYGKYSFAFELDTGYGFLPARQLP 396

RESULT 13
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
R;Matthews, L.
C;Accession: T20507; T24289
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19284

```

A:Accession: T20507
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-323 <W1>
A:Cross-references: EMBL:Z78411; PIDN:CA801647.1; GSPDB:GN00023; CESP:F02D8.4
A:Experimental source: clone F02D8
R:Wild, A.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19869
A:Accession: T24289
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-323 <W12>
A:Cross-references: EMBL:Z78413; PIDN:CA801666.1; GSPDB:GN00023; CESP:F02D8.4
A:Experimental source: clone T01C3
C:Genetics:
A:Gene: CESP:F02D8.4
A:Map position: 5
A:Introns: 33/3; 67/3; 201/3; 228/1; 257/2
C:Superfamily: carboxypeptidase

Query Match          30.1%; Score 403; DB 2; Length 323;
Best Local Similarity 33.5%; Pred. No. 1.3e-28;
Matches 89; Conservative 52; Mismatches 81; Indels 44; Gaps 7;

Qy 18 WIEFITERHPDMLTKIHGSSFEKYPIYLVKVSKEQTAKNAIWDCGIIHAREWISPAFC 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 FLNSLAQQYNDVKLVQIGNSYEGRSITAVRIA-DCGSSKPIVWIDAGIHAREWISYVNA 61

Qy 78 LWFT-----CHRMWTKRNSFYVNHHCITDIL 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 LYLITIVSQPAYRNLLDSVLQVVPVNTNPDGYEYRTNDRMWKTRSFTRNSRCAGADA 121

Qy 105 NRPFSKXHCWCEGASSSCSETYCGLYPESEPEVKAVASFLRRNINOIKAYISMHSYSQH 164
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 NRNVYP-YWGTQGVSHSCSEIFCGSPQSEPEVLATNAIIRDEERIKGYIALHSYQOE 180

Qy 165 IVPYYSVT-RSKSKDHBELSLVA---SEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDD 220
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 ILVPWGHGTQRTYPTDVQDLIQVGRAMASAIRAV---NNTDYTEVNSGDLGYPAAGASDD 236

Qy 221 WIYDLGIKYSFTSN-PPVEKLLPLSL 245
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 WAKSRGIKYSYTIELSPIDDFTGFSL 262

RESULT 14
T3327
hypothetical protein T06A4.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C:Accession: T33527
R:Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid T06A4.
A:Reference number: Z21364
A:Accession: T33527
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-528 <WUX>
A:Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A:Experimental source: strain Bristol N2; clone T06A4
C:Genetics:
A:Gene: CESP:T06A4.1
A:Map position: 1
A:Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C:Superfamily: carboxypeptidase

Query Match          29.6%; Score 396.5; DB 2; Length 528;
Best Local Similarity 35.5%; Pred. No. 9e-28;
Matches 98; Conservative 43; Mismatches 80; Indels 55; Gaps 10;

Qy 6 YEQVHSNLEYISWTEFITERHPDMLTKIHGSSPEKYPVLYLVKVSKEQT--AKNAIWD 63

```

Db 140 YDKYNSLEBIEQTEMRLKKEYTMTLIDIGOSHENRTLLVMKITGRKNPLUGSKISMWID 199
Qy 64 CQIHAREWISPAFCLEWFIH-----NR 85
Db 200 AGIHAREWIAPATAM-YTAHELLLGYENDATVAKLMDHIDFYLPVMPNPDGYEYSREKQR 258
Qy 86 MWRKNRS-----FYANNHCITGDLNRNPFASKHWCBEAGASSSCSCTCYGLYPESEPEV 138
Db 259 MWRKNRSAPAKCARQFTSTVCCSGVDLNRNF-DWFMASTGSSDPCHDTYHGSAAFSEPE 317
Qy 139 KAVASFLRNINQIKAYISMHSYSQIHVPYPSYTR-SKSKD-HEELSLVASEAVRAIKET 196
Db 318 QAVRDFLEQNTPE--AFISLHYSQWMLPYGHRKQSYQDQYHTGLRPLRALRATKALYEL 375
Qy 197 SKNTRYTHGHGSETLYLAPGGDDWIY-DLGIKYSF 231
Db 376 -YGTYQVGTGADLMEYASGGSHDWAKGQKVPIAY 410

RESULT 15
CPCYB
carboxypeptidase B (EC 3.4.17.2) - broad-fingered crayfish
C:Species: Astacus astacus, Astacus fluviatilis (broad-fingered crayfish)
C:Date: 30-Jun-1988 #sequence revision 30-Jun-1988 #text_change 06-Dec-1996
C:Accession: A05141; A14178
R:Titani, K.; Ericsson, L.H.; Kumar, S.; Jakob, F.; Neurath, H.; Zwilling, R.
Biochemistry 23, 1245-1250, 1984
A:Title: Amino acid sequence of crayfish (Astacus fluviatilis) carboxypeptidase B.
A:Reference number: A90482
A:Accession: A05141
A:Molecule type: protein
A:Residues: 1-303 <Tit>
C:Superfamily: carboxypeptidase
C:Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F:241,264/Active site: Tyr, Glu #status predicted

Query Match 28.1%; Score 375.5; DB 1; Length 303;
Best Local Similarity 33.1%; Pred. No. 3.5e-26;
Matches 87; Conservative 45; Mismatches 92; Indels 39; Gaps 6;
Qy 6 YEQTHSLNEIYSWIFETERHPDMLTKIHGGSFEKYPLYVLKVSKEQTAKNAIWDG 65
Db 3 WTSYHDYDEINAWLDSLATYDELASVEDVGLSYEGRTMKLLKL-GRGGADKPIFIDGG 61
Qy 66 IHAREWISPAFCLEWFIH-----NRMWRKNR 91
Db 62 IHAREWIAPTVTYIVNPFVSNATYDDILSNVNFVMTINPDGYAYTFTDDLWRKTR 121
Qy 92 SFYAN-NHCIGTDLNRNPFASKHWCBEAGASSSCSCTCYGLYPESEPEVKVASFLLRNIN 150
Db 122 SETGSVLGCKGADPNRW-SFHWDEVGASDSCPSDIYAGPEPFSEVENVRDQILEYAA 180
Qy 151 QIKAYISMHSYSQIHVPYPSYTRSKDHEELSLVASEAVRAIKETSKNTRYTHGHGSET 210
Db 181 NIKVYLTFHYSQWMLPYGFTSDLPDDWQDLDTLATNAVDALTAV-HGTRYEGSSTNT 239
Qy 211 LYLAPGGDDWIY-DLGIKYSF 232
Db 240 IYAAAGSGDDWAKGEGGVKIAYT 262

Search completed: January 1, 2004, 19:34:33
Job time : 19.1978 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:03:36 ; Search time 9.75783 Seconds
(without alignments)
1185.568 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQYHSLNEIYSWIE.....IKYSFTSNPPVKLLPLSLK 246

Scoring table: BLOSUM62.
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	601	44.9	415	1	P19223 rattus norv
2	586	43.8	306	1	P00732 bos taurus
3	578	43.2	416	1	P55261 canis famil
4	575	43.0	401	1	P09955 sus scrofa
5	567	42.4	417	1	P15086 homo sapien
6	561	41.9	417	1	P15089 mus musculu
7	552	41.3	309	1	P21961 rattus norv
8	552	41.3	417	1	P15088 homo sapien
9	502.5	37.6	419	1	P00730 bos taurus
10	488.5	36.5	419	1	P00731 rattus norv
11	483	36.1	417	1	P19222 rattus norv
12	465.5	34.8	419	1	P15085 homo sapien
13	459	34.3	417	1	P48052 homo sapien
14	458	34.2	421	1	Q9ui42 homo sapien
15	411.5	30.8	304	1	P42788 simulum v1
16	381.5	28.5	433	1	O02350 anopheles g
17	375.5	28.1	303	1	P04069 astacus flu
18	286.5	21.4	424	1	P29068 thermoaactin
19	286.5	21.4	430	1	P38836 saccharomyc
20	282.5	21.1	451	1	F18143 streptomyc
21	255	19.1	434	1	P39041 streptomyc
22	118.5	8.9	1377	1	O89001 mus musculu
23	115.5	8.6	1380	1	O75976 homo sapien
24	103.5	7.7	445	1	P70080 gallus gall
25	103.5	7.7	734	1	Q96sm3 homo sapien
26	99	7.4	444	1	P17752 homo sapien
27	98	7.3	722	1	Q92100 mus musculu
28	97	7.2	297	1	Q15006 homo sapien
29	96.5	7.2	376	1	P54497 bacillus su
30	95	7.1	444	1	P17290 oryctolagus
31	95	7.1	707	1	P54789 schizosacch
32	91	6.8	262	1	P51983 escherichia
33	90.5	6.8	434	1	P04836 bos taurus

34	90.5	6.8	476	1	CBPB_RAT
35	89.5	6.7	476	1	CBPH_MOUSE
36	89	6.7	481	1	TPH_XENLA
37	88.5	6.6	476	1	CBPH_HUMAN
38	86	6.4	382	1	HMBP_DROME
39	86	6.4	447	1	TPH1_MOUSE
40	85.5	6.4	465	1	STHA_ECO57
41	85.5	6.4	465	1	STHA_ECOLI
42	85.5	6.4	1406	1	CBPD_DROME
43	85	6.4	1255	1	PER2_HUMAN
44	83.5	6.2	319	1	UN30_CABEL
45	83.5	6.2	444	1	TPH1_RAT

P15087	rattus norv
Q00493	mus musculu
Q92142	xenopus lae
P16870	homo sapien
P22809	drosophila
P17532	mus musculu
Q8x727	escherichia
P27306	escherichia
P42787	drosophila
O15055	homo sapien
P52906	caenorhabdi
P09810	rattus norv

ALIGNMENTS

RESULT 1

ID	CBPB_RAT	STANDARD;	PRT;	415 AA.
AC	P19223;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Carboxypeptidase B precursor (EC 3.4.17.2).			
GN	CPB.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=101116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=89034324; PubMed=3182872;			
RX	Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;			
RT	"Structural characterization of the rat carboxypeptidase A1 and B			
RT	genes. Comparative analysis of the rat carboxypeptidase gene			
RT	family.";			
RL	J. Biol. Chem. 263:17837-17845(1988).			
CC	-I- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =			
CC	peptide + L-lysine(or L-arginine).			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M23959; AAA40872.1; .			
DR	EMBL; M23947; AAA40872.1; JOINED.			
DR	EMBL; M23950; AAA40872.1; JOINED.			
DR	EMBL; M23952; AAA40872.1; JOINED.			
DR	EMBL; M23953; AAA40872.1; JOINED.			
DR	EMBL; M23954; AAA40872.1; JOINED.			
DR	PIR; A32129; A32129.			
DR	HSSP; P09955; INSA.			
DR	MEROPS; M14.003; .			
DR	InterPro; IPR003146; Propep_M14.			
DR	InterPro; IPR00834; Zn_carboPept.			
DR	Pfam; PF02244; Propep_M14; 1.			
DR	Pfam; PF00246; Zn_carboPept; 1.			
DR	SMART; SM00631; Zn_pept; 1.			
DR	PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.			
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.			
KW	Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.			
FT	SIGNAL 1 13			
FT	CHAIN 109 415			
FT	PROPEP 14 108			
FT	ACTIVATION PEPTIDE B.			
FT	CARBOXYPEPTIDASE B.			
FT	ZINC (BY SIMILARITY).			
FT	ZINC 174 177			
FT	ZINC (BY SIMILARITY).			
FT	ZINC 177 177			
FT	ZINC (BY SIMILARITY).			
FT	ZINC 302 302			
FT	ZINC (BY SIMILARITY).			

FT ACT SITE 376 376 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 171 184 BY SIMILARITY.
 FT DISULFID 243 266 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 SQ SEQUENCE 415 AA; 47515 MW; 8EA06CCADE30B6F2 CRC64;
 Query Match 44.9%; Score 601; DB 1; Length 415;
 Best Local Similarity 44.6%; Pred. No. 2,7e-48;
 Matches 120; Conservative 42; Mismatches 67; Indels 40; Gaps 5;
 QY 1 ASASYEQYHSLNEIWSIEFITERHDPMLTKIHGSSFEKYPYLVKVGKQAKNAI 60
 DB 109 ASGHSYTKYNKWTETIAWIOQVATDNEDLVTSQVIGTFEGRNMYLKI-GKTRPNKPAI 167
 QY 61 WIDCGTHAREWISPAFLWFI-----GH-----83
 DB 168 FIDCGFHAREWISPAFLWFI-----GH-----83
 QY 84 NRMWRKRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVAS 143
 DB 228 DRWKRTRSTWAGSSCLGVPRNRF-NAGCEVGASRSPSETYCGPAPSEKETKALAD 286
 QY 144 FLRNINQIKAYISMSYSQHIYVFPYSYTRSKSDHSELSVASEAVRAIEKTSKNTRYT 203
 DB 287 FIRNLSSTIKAYLTIHSYQMLYPYSDYKLPENYEELNALVKAAREL-ATLHGKTYT 345
 QY 204 HGHSELYLAPGGDDWIYDLGIKYSFT 232
 DB 346 YGPGATTIYPAGGDDWSYDQGIKYSFT 374

RESULT 2

CBPB_BOVIN STANDARD; PRT; 306 AA.
 ID_CBPB_BOVIN
 AC PF0732;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase B (EC 3.4.17.2).
 GN CFB.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=75217824; PubMed=1057162;
 RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
 RT "Amino-acid sequence of bovine carboxypeptidase B";
 RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670 (1975).
 RN [2]
 RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
 RX MEDLINE=74260705; PubMed=4833744;
 RA Schmidt J.J., Hirs C.H.W.;
 RT "Primary structure of bovine carboxypeptidase B. Inferences from the
 RT locations of the half-cystines and identification of the active site
 RT arginine.";
 RL J. Biol. Chem. 249:3756-3764 (1974).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
 RX MEDLINE=76265065; PubMed=957425;
 RA Schmid M.F., Herriott J.R.;
 RT "Structure of carboxypeptidase B at 2.8-A resolution.";
 RL J. Mol. Biol. 103:175-190 (1976).
 RN [4]
 RP ACTIVE SITE.
 RX MEDLINE=70007159; PubMed=5344132;
 RA Plummer T.H. Jr.;
 RT "Isolation and sequence of peptides at the active center of bovine
 RT carboxypeptidase B.";
 RL J. Biol. Chem. 244:5246-5253 (1969).
 RN [5]

RP ACTIVE SITE.
 RX MEDLINE=73061487; PubMed=4565668;
 RA Kimmel M.T., Plummer T.H. Jr.;
 RT "Identification of a glutamic acid at the active center of bovine
 RT carboxypeptidase B.";
 RL J. Biol. Chem. 247:7864-7869 (1972).
 CC -I- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =
 CC Peptide + L-lysine (or L-arginine).
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
 DR PIR; A93797; CPB0B.
 DR PDB; 1CPB; 30-SEP-83.
 DR MEROPS; M14.003; -.
 DR InterPro; IPR000834; Zn carbOpept.
 DR Pfam; PF00246; Zn carbOpept; 1.
 DR SMART; SM00631; Zn_pept; 1.
 DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
 DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
 KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
 FT DISULFID 63 76
 FT DISULFID 135 158
 FT DISULFID 149 163
 FT METAL 66 66 ZINC.
 FT METAL 69 69 ZINC.
 FT METAL 194 194 ZINC.
 FT ACT_SITE 246 246
 FT ACT_SITE 268 268 NUCLEOPHILE.
 SQ SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;
 Query Match 43.8%; Score 586; DB 1; Length 306;
 Best Local Similarity 43.9%; Pred. NO. 4.5e-47;
 Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;
 QY 6 YEQYHSLNEIWSIEFITERHDPMLTKIHGSSFEKYPYLVKVGKQAKNAIWDG 65
 DB 6 YEKYNWETIEAWTEQVASENPDLSRSAIGTTFGLNTIYLLKV-GKPGSNKPAVFMDCG 64
 QY 66 IHAREWISPAFLWFI-----GH-----NRWR 88
 DB 65 FHAREWISPAFLWFI-----GH-----NRWR 124
 QY 89 KRSFYANNHCIGTDLNRNFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 148
 DB 125 KTRSTRAGSSCTGTLNRNF-DAGWCISGANNPCSETYCGSAASESEKSKAVDFIRNH 183
 QY 149 INQIKAYISMSYSQHIYVFPYSYTRSKSDHSELSVASEAVRAIEKTSKNTRYTHGHS 208
 DB 184 LSSIKAYLTIHSYQMLYPYSDYKLPKNVVELNLAKGAVKKL-ASLHGTTYSYPGA 242
 QY 209 ETLYLAPGGDDWIYDLGIKYSFT 232
 DB 243 TTIYPASGSDWAYDQGIKYSFT 266
 RESULT 3
 CBPB_CANFA STANDARD; PRT; 416 AA.
 ID_CBPB_CANFA
 AC PS5261;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
 DE membrane associated protein) (ZAP47).
 GN CPB1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Fukuoka S.-I.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -I- CATALYTIC ACTIVITY: Peptidyl-L-Lysine (or L-arginine) + H(2)O =

FT	STRAND	11	11
FT	HELIX	20	32
FT	STRAND	36	39
FT	HELIX	43	45
FT	STRAND	50	56
FT	HELIX	58	60
FT	HELIX	61	69
FT	TURN	70	72
FT	STRAND	74	79
FT	HELIX	82	87
FT	TURN	88	88
FT	TURN	101	102
FT	HELIX	107	120
FT	STRAND	122	124
FT	STRAND	125	132
FT	TURN	134	135
FT	STRAND	138	144
FT	TURN	149	150
FT	STRAND	153	157
FT	TURN	162	163
FT	HELIX	165	181
FT	TURN	182	184
FT	HELIX	186	194
FT	STRAND	196	200
FT	HELIX	205	213
FT	TURN	214	214
FT	TURN	216	217
FT	STRAND	224	224
FT	TURN	226	227
FT	STRAND	231	231
FT	HELIX	235	237
FT	TURN	243	244
FT	STRAND	249	249
FT	TURN	252	253
FT	TURN	255	256
FT	STRAND	257	257
FT	TURN	262	263
FT	HELIX	266	277
FT	TURN	278	281
FT	STRAND	282	289
FT	STRAND	294	297
FT	TURN	307	308
FT	HELIX	309	327
FT	STRAND	332	335
FT	HELIX	336	339
FT	HELIX	347	353
FT	TURN	354	355
FT	STRAND	359	363
FT	TURN	371	372
FT	HELIX	376	378
FT	HELIX	379	398
FT	TURN	399	401
SO	SEQUENCE	401 AA;	451

Qy 209 ETILYAPGGDDWIYDLGIKYSFT 232
| : | | | | | | | | | |
Db 338 TTITPAAGGSDWAYNQGIKYSFT 361

RESULT 5
CEBP_HUMAN STANDARD; PRT; 417 AA.
AC PI5086; O60834; Q96BQ8;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase B precursor [EC 3.4.17.2] (Pancreas-specific protein
DE (PASP)).
GN CPB1 OR CPB OR PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC TISSUE=Pancreas;
RX MEDLINE=92129345; PubMed=1370825;
RA Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA French C.K.;
RT "Isolation of a cDNA encoding a human serum marker for acute
RT pancreatitis. Identification of pancreas-specific protein as
RT pancreatic procarboxypeptidase B.";
RL J. Biol. Chem. 267:2575-2581(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=98182241; PubMed=9524066;
RA Alcy P., Catasus L., Villegas V., Reverter D., Vendrell J.,
RA Aviles F.X.;
RT "Comparative analysis of the sequences and three-dimensional models
RT of human procarboxypeptidases A1, A2 and B";
RL Biol. Chem. 379:149-155(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altshul S.F., Zesberg B., Buecow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toehynki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Vallalon D.K., Musny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.N., Krzywinski M.T., Skalska U., Smalhus D.E.,
RA Schnorch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 16-43.
RC TISSUE=Pancreas;
RX MEDLINE=89153096; PubMed=2920728;
RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT "Purification and properties of five different forms of human
RT procarboxypeptidases.";
RL Eur. J. Biochem. 179:609-616(1989).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2O) =
CC peptide + L-lysine(or L-arginine).

RA Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT "Human mast cell carboxypeptidase. Purification and
RL Characterization.";
CC J. Clin. Invest. 83:1630-1636 (1989).
CC -I- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H₂O = peptide + L-
CC amino acid.
CC -I- SUBCELLULAR LOCATION: Secretory granules.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M2717; AAA35652.1; -;
DR EMBL; M73720; AAA59568.1; -;
DR EMBL; M73716; AAA59568.1; JOINED.
DR EMBL; M73717; AAA59568.1; JOINED.
DR EMBL; M73718; AAA59568.1; JOINED.
DR EMBL; M73719; AAA59568.1; JOINED.
DR EMBL; BC012613; AAH12613.1; -;
DR EMBL; S40234; AAB22578.2; ALT_SEQ.
DR EMBL; A43929; A43929.
DR HSSP; P09955; INSA.
DR MEROPS; M14.010; -;
DR Genew; HGNC:2298; CPA3.
DR MIM; 114851; -;
DR GO; GO:0030141; C:secretory granule; NAS.
DR GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn carboxypeptidase.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn carboxypeptidase; 1.
DR SMART; SM00631; Zn pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT ZN 1; 1.
DR PROSITE; PS00133; CARBOXYPEPT ZN 2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 15
FT PROPEP 16 109 ACTIVATION PEPTIDE.
FT CHAIN 110 417 MAST CELL CARBOXYPEPTIDASE A.
FT METAL 176 176 ZINC (BY SIMILARITY).
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 304 304 ZINC (BY SIMILARITY).
FT ACT_SITE 378 378 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 245 268 BY SIMILARITY.
FT CONFLICT 63 63 N -> K (IN REF. 3).
SQ SEQUENCE 417 AA; 48700 MW; 8CB90DB758117B24 CRC64;

Query Match 41.3%; Score 552; DB 1; Length 417;
Best Local Similarity 41.4%; Pred. No. 9.6e-44;
Matches 109; Conservative 45; Mismatches 69; Indels 40; Gaps 5;

QY 6 YE0YHSLNETYSWIEFTTERHPDMLTKIHGSGFEKYPVLVYKVGKEQTAKNAIWDG 65
DB 116 YAKNNWEKIVATERMDKYPENVRKIKGSTVEDNPLVLKI-GEKERRKAI FMDG 174

QY 66 IHAREWISPAFLWFI-----GH-----NRMWR 88
DB 175 IHAREWVSPAFQWVQVATKYGRNKMTPKLLDRMNFYLPVFNVDGYIWSWTKRMWR 234

QY 89 KNRSFVANHICITDLNRNPASKHWCBEAGSSSCSETYCGLYPESPEVKAVASFLRN 148
DB 235 KNRKQNSKICITDLNRNP-NASWNSIPNTNDPCADNYRGSAPSEKETKAVTNFIRSH 293

QY 149 INOKIYVISHMSYSHQIVRPVPSYTRSKSKDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
DB 294 LNEIKVITTHSYQSMLLPPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRIYGPTE 352

QY 209 ETLYLAPGGDDWIYDLGIKYSF 231
DB 353 STIYPISSSLDWAYDLGIKHTF 375

RESULT 9
CBPA_BOVIN STANDARD; PRT; 419 AA.
ID CBPA_BOVIN
AC P00730;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase A precursor (EC 3.4.17.1).
GN CPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=91151335; PubMed=1998496;
RA le Hueeou I., Guilloreau P., Toulliec R., Puigserver A., Wicker C.;
RT "Cloning and nucleotide sequence of a bovine pancreatic
RT preprocarboxypeptidase A cDNA.";
RL Biochem. Biophys. Res. Commun. 175:110-116 (1991).
RN [2]
RS SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RX MEDLINE=96096549; PubMed=8522204;
RA Goo J.H., Kim K.H., Choi K.Y.;
RT "Cloning, sequencing and expression of the gene encoding a major
RT allotypic preprocarboxypeptidase A from bovine pancreas.";
RL Gene 165:333-334 (1995).
RN [3]
RS SEQUENCE OF 111-417
RX MEDLINE=71155183; PubMed=5102489;
RA Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RT chymotryptic peptides of the cyanogen bromide fragment P-I.";
RL Biochemistry 10:938-950 (1971).
RN [4]
RS REVISIONS TO 138 AND 141.
RX MEDLINE=72138789; PubMed=5143102;
RA Petra P.H., Hermodson M.A., Walsh K.A., Neurath H.;
RT "Characterization of bovine carboxypeptidase A (Allan).";
RL Biochemistry 10:4023-4025 (1971).
RN [5]
RS SEQUENCE OF 17-120.
RX MEDLINE=89150306; PubMed=3147705;
RA Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neurath H.;
RT "The amino acid sequence of the activation peptide of bovine pro-
RT carboxypeptidase A.";
RL Biochimie 70:1137-1142 (1988).
RN [6]
RS X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
RX MEDLINE=83294519; PubMed=6887246;
RA Rees D.C., Lewis M., Lipscomb W.N.;
RT "Refined crystal structure of carboxypeptidase A at 1.54-A
RT resolution.";
RL J. Mol. Biol. 168:367-387 (1983).
RN [7]
RS X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX MEDLINE=96003618; PubMed=7556081;
RA Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT "The three-dimensional structure of the native ternary complex of
RT bovine pancreatic procarboxypeptidase A with proproteinase E and
RT chymotrypsinogen C.";
RL EMBO J. 14:4387-4394 (1995).
RN [8]
RS VARIANT ALLELIC.
RX MEDLINE=69283620; PubMed=5817619;
RA Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;

AC P00731;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN CPAL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82105986; PubMed=6275388;
RA Quinco C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
RA Pictet R.L., Valenzuela P., Rutter W.J.;
RT "Rat procarboxypeptidase A: cDNA sequence and preliminary
RT characterization of the gene."
RN Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family."
RL J. Biol. Chem. 263:17837-17845(1988).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; V01232; CAA24542.1; -;
DR EMBL; J00713; AAA40893.1; -;
DR EMBL; M23990; AAA40955.1; -;
DR EMBL; M23960; AAA40955.1; JOINED.
DR EMBL; M23985; AAA40955.1; JOINED.
DR EMBL; M23986; AAA40955.1; JOINED.
DR EMBL; M23987; AAA40955.1; JOINED.
DR EMBL; M23988; AAA40955.1; JOINED.
DR EMBL; M23989; AAA40955.1; JOINED.
DR PIR; A00911; CPRTA.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.001; -;
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carboxept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carboxept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 110 ACTIVATION PEPTIDE.
FT CHAIN 111 419 CARBOXYPEPTIDASE A1.
FT METAL 179 179 ZINC (BY SIMILARITY).
FT METAL 182 182 ZINC (BY SIMILARITY).
FT METAL 306 306 ZINC (BY SIMILARITY).
FT ACT_SITE 358 358 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 380 380 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 248 271 BY SIMILARITY.
FT CONFLICT 196 196 I -> V (IN REF. 1).
FT CONFLICT 261 263 FGM -> LGK (IN REF. 1).
FT CONFLICT 347 347 K -> E (IN REF. 1).
SQ SEQUENCE 419 AA; 47197 MW; BB002D1CB99B7491 CRC64;

Query Match 36.5%; Score 488.5; DB 1; Length 419;
Best Local Similarity 39.4%; Pred. No. 7.6e-38;
Matches 104; Conservative 42; Mismatches 77; Indels 41; Gaps 5;
QY 6 YEQYHSLNEIYSWIEPITERHPDMLTKIHGSGFEKYPYLVKVGKQETAKNAIWDG 65
DB 119 YATYHTLDIYEFMDLLVAEHPQLVSKIQIGNTFEGRPPIHLKFS-TGGTNRRAIMDTG 177
QY 66 IHAREWISPAFLWFIGH-----NRMWR 88
DB 178 IHSREWVTQASGVWFPAKKITKDYGDPTFTAVLDNMDIFLEIVTNPDPGAYTHKTNRMWR 237
QY 89 KRSFYANNHCITGLNRNPFASKHWCSEGASSSCSETYCGLYPESEPEVKAVASFLRRN 148
DB 238 KTRSHQTGSLCVGVDPNRMW-DAGFGMAGASSNPCSETYRGKFPNSEVEVKISVDFTVSH 296
QY 149 INQIKAYISMHSYSHIVFPYPSYTRSKSDHSELSLVASEAVRAIEKTSKNTYTHGHGS 208
DB 297 GN-IAFISIHYSQLLLYPGYGTSEPAQDQLAKSAVTAL-TSLHGTKEYGSGII 354
QY 209 ETLYLAPGGDDWIYDLGIKYSFT 232
DB 355 DTIYQASGTIDWTYSQGIKYSFT 378
RESULT 11
ID_CBP2_RAT STANDARD; PRT; 417 AA.
AC P1922;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN CPA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas.
RX MEDLINE=89034323; PubMed=3182871;
RA Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
RA Graf M., Rutter W.J.;
RT "A novel rat carboxypeptidase, CPA2: characterization, molecular
RT cloning, and evolutionary implications on substrate specificity in
RT the carboxypeptidase gene family."
RL J. Biol. Chem. 263:17828-17836(1988).
RN [2]
RP SEQUENCE OF 131-143 FROM N.A.
RX MEDLINE=95386501; PubMed=7657630;
RA Normant E., Gros C., Schwartz J.C.;
RT "Carboxypeptidase A isoforms produced by distinct genes or
RT alternative splicing in brain and other extrapancreatic tissues."
RL J. Biol. Chem. 270:20543-20549(1995).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=92105124; PubMed=1761558;
RA Faming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
RT "Structural evolution of an enzyme specificity. The structure of rat
RT carboxypeptidase A2 at 1.9-A resolution."
RL J. Biol. Chem. 266:24606-24612(1991).
CC -1- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC residues.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC

or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; M23721; AAA40956.1; -.
CC EMBL; M23714; AAA40956.1; JOINED.
CC EMBL; M23715; AAA40956.1; JOINED.
CC EMBL; M23716; AAA40956.1; JOINED.
CC EMBL; M23717; AAA40956.1; JOINED.
CC EMBL; M23718; AAA40956.1; JOINED.
CC EMBL; M23719; AAA40956.1; JOINED.
CC EMBL; M23720; AAA40956.1; JOINED.
CC EMBL; S79837; -. NOT_ANNOTATED_CDS.
CC PIR; A32128; A32128.
CC HSSP; P48052; IAYE.
CC MEROPS; M14.002; -.
CC InterPro; IPR003146; Propep_M14.
CC InterPro; IPR000834; Zn_carboxypept.
CC Pfam; PF02244; Propep_M14; 1.
CC Pfam; PF00246; Zn_carboxypept; 1.
CC PRINTS; PR00765; CRBOXYPTASEA.
CC SMART; SM00631; Zn_pept; 1.
CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
CC Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
CC SIGNAL 1 16
CC PROPEP 17 112
CC CHAIN 113 417
CC METAL 177 177
CC METAL 180 180
CC METAL 304 304
CC ACT_SITE 378 378
CC DISULFID 246 269
CC FT DISULFID 318 352
CC FT SEQUENCE 417 AA; 46912 MW; BEDCC41A830F2D45 CRC64;

Query Match 36.1%; Score 483; DB 1; Length 417;
Best Local Similarity 38.2%; Pred. No. 2.5e-37;
Matches 109; Conservative 47; Mismatches 75; Indels 54; Gaps 9;

OY 6 YEYHSLNELYSWIEFITERHPDMLTKIHGSSPEKYPVLYLVKVS_CKEQTAKNAWIDC 64
DB 118 FEAYHTLEEIQEMDLVAENPLGLSVKNLGSFENRPMNVLKFTSGD---KPAIWLDA 174
OY 65 GIAREWISPAFLW---FI-----GH-----NRMW 87
DB 175 GIAREWVTOATLWNTANKIASDYGTPAITSLLNTLIDIFLLPWTNPDPGVVFQTTNRMW 234
OY 88 RKRFSYANNHCIGTDLNRNFASKHWCCEGASSSCSCTCYGLYPESEPEPKAVAFLLR 147
DB 235 RKTRSKRSGCGVGVDPNRNW-DANFGCGGASSSPCSDSYHGPKNSEVEVKSIQVFIKS 293
OY 148 NIQIKAYISWHSYSQIHVFPYSYTRSKSKDHBELSLVASENVAETKSKNTRYTHGCG 207
DB 294 H-GKVKAFITLHSYSQMLPPYGYKCTKDPDFNDELVAQAQAAL-KRLHGTSYKVGPI 351
OY 208 SETLYLAPGGDDWIYDLGKIKYFTSN-----PPVEKLLP 242
DB 352 CSVIYQASGSGSIDWAYDLGKIKYFAFLRDTAFYGLLPKAKIUP 396
```

RESULT 12

```
CBPI_HUMAN
ID CBPI_HUMAN STANDARD; PRT; 419 AA.
AC P15085; Q9BS67;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN CPA1 OR CPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=93038569; PubMed=1417781;
RA Catasus L., Villegas V., Pascual R., Aviles F.X.,
RA Wicker-Planquart C., Puigserver A.;
RT "cDNA cloning and sequence analysis of human pancreatic
RL procarboxypeptidase A1.";
RN Biochem. J. 287:299-303(1992).
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE OF 17-42.
RX MEDLINE=89153096; PubMed=2920728;
RA Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT "Purification and properties of five different forms of human
RL procarboxypeptidases.";
CC Eur. J. Biochem. 179:609-616(1989).
CC -1- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X67318; CAA47732.1; -.
CC EMBL; BC005279; AAH05279.1; -.
CC PIR; S29127; S29127.
CC HSSP; P00730; 1PYT.
CC MEROPS; M14.001; -.
CC Genew; HGNC:2296; CPA1.
CC MIM; 114850; -.
CC GO; GO:0004182; P:carboxypeptidase A activity; TAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
CC InterPro; IPR003146; Propep_M14.
CC InterPro; IPR000834; Zn_carboxypept.
CC Pfam; PF02244; Propep_M14; 1.
CC Pfam; PF00246; Zn_carboxypept; 1.
CC PRINTS; PR00765; CRBOXYPTASEA.
CC SMART; SM00631; Zn_pept; 1.
CC PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
CC PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
CC Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
KW SIGNAL 1 16
KW PROPEP 17 110
KW CHAIN 111 419
FT PROPEP 111 419
FT CHAIN 111 419
FT CARBOXYPEPTIDASE A1.
```

FT METAL 179 179 ZINC (BY SIMILARITY).

FT METAL 182 182 ZINC (BY SIMILARITY).

FT METAL 306 306 ZINC (BY SIMILARITY).

FT ACT SITE 358 358 PROTON DONOR (BY SIMILARITY).

FT ACT SITE 380 380 NUCLEOPHILE (BY SIMILARITY).

FT DISULFID 248 271 BY SIMILARITY.

FT CONFLICT 276 276 H -> R (IN REF. 2).

SQ SEQUENCE 419 AA; 47140 MW; 439FAFFAE958B1 CRC64;

Query Match 34.8%; Score 465.5; DB 1; Length 419;

Best Local Similarity 37.9%; Pred. No. 1e-35;

Matches 100; Conservative 44; Mismatches 79; Indels 41; Gaps 5;

QY 6 YEOYHSLNELYSWIEFTEHPDMLTKIHGSSFEKYPYLVKVSKEQTAKNAIWDG 65

DB 119 YATYHTLEIYDLDLVAENPLVSKIQIGNTYEGRIYVLKFS-TGGSKRPAINIDTG 177

QY 66 IHAREWISPAFLCWF-----GHRNWR 88

DB 178 IHSREWTOAGVWFPAKKITQDYGQDAFTALDITDIFLEIVNPDGFAFTHSTNRMR 237

QY 89 KNRSPYANNHCITDLNRNPFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFLRN 148

DB 238 KTRSHTAGSLICGVDPNRNDAGFGL-SCASSNPCSETYHGKFNSEVEKSVDFVKOH 296

QY 149 INQIKAYISMHSYQHVFPYSTRSKSDHBEHLSLVAENPRAIKTSKNTRYTHGHS 208

DB 297 GN-IKAFISHTSYQLMTPYGYKTEPVPDQDELQSLAAVTAL-ASLYGTFKFNYSII 354

QY 209 ETLYLAPGGGDWYLDGIKYSET 232

DB 355 KAIYQASGSTIDWTSQGIKYSET 378

RESULT 13

CPB2_HUMAN STANDARD; PRT; 417 AA.

AC P48052; Q96A12; Q96QN3;

DT 01-FEB-1996 (Rel. 33, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Carboxypeptidase A2 precursor (EC 3.4.17.15).

GN CPA2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.

RC TISSUE=Pancreas;

RX MEDLINE=95204457; PubMed=7896805;

RA Catasus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,

RA Billeter M.,

RT "The sequence and conformation of human pancreatic

RT procarboxypeptidase A2. cDNA cloning, sequence analysis, and

RT three-dimensional model.",

RL J. Biol. Chem. 270:6651-6657 (1995).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain, and Pancreas;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,

RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,

RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.",

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).

RX MEDLINE=98111000; PubMed=9450539;

RA Reverter D., Garcia-Saez I., Catasus L., Vendrell J., Coll M.,

RA Aviles F.X.,

RT "Characterisation and preliminary X-ray diffraction analysis of human

RT pancreatic procarboxypeptidase A2.",

RL FEBS Lett. 420:7-10 (1997).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).

RX MEDLINE=98046021; PubMed=9384570;

RA Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.,

RT "The three-dimensional structure of human procarboxypeptidase A2.

RT Deciphering the basis of the inhibition, activation and intrinsic

RT activity of the zymogen.",

RL EMBO J. 16:6906-6913 (1997).

CC -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A

CC (EC 3.4.17.1), but with a preference for bulkier C-terminal

CC residues.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U19977; AAA74425.1; -

DR EMBL; BC007009; AAH07009.1; -

DR EMBL; BC014571; AAH14571.1; -

DR EMBL; BC015140; AAH15140.1; -

DR PIR; A56171; A56171.

DR PDB; 1AYE; 13-JAN-99.

DR PDB; 1D7D; 12-JUL-00.

DR PDB; 1O6X; 30-JAN-03.

DR MEROPS; M14.002; -

DR Genew; HGNC:2297; CPA2.

DR MIM; 606688; -

DR GO; GO:0004182; P:carboxypeptidase A activity; TAS.

DR GO; GO:0007039; P:vacuolar protein catabolism; TAS.

DR InterPro; IPR003146; Propep M14.

DR InterPro; IPR000834; Zn_carBopept.

DR Pfam; PF02244; Propep M14; 1.

DR Pfam; PF00246; Zn_carBopept; 1.

DR PRINTS; PR00765; CRBOXYPTASEA.

DR SMART; SM00631; Zn_pept; 1.

DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.

DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.

KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;

KW 3D-structure.

FT SIGNAL 1 16 POTENTIAL.

FT PROPEP 17 112 ACTIVATION PEPTIDE.

FT CHAIN 113 417 CARBOXYPEPTIDASE A2.

FT METAL 177 177 ZINC.

FT METAL 180 180 ZINC.

FT METAL 304 304 ZINC.

FT ACT SITE 378 378 NUCLEOPHILE.

FT DISULFID 246 269

FT DISULFID 318 352

FT CONFLICT 80 80 E -> G (IN REF. 2; AAH07009).

FT CONFLICT 302 302 T -> I (IN REF. 1).

SQ SEQUENCE 417 AA; 46828 MW; 3D9C073A98700A83 CRC64;

```
Query Match 34.3%; Score 459; DB 1; Length 417;
Best Local Similarity 36.5%; Pred. No. 4.2e-35;
Matches 103; Conservative 46; Mismatches 79; Indels 54; Gaps 7;

QY 9 YHSLNLIYSWIEFITERHPDMLTKIHIGSFKEYPLVVLKVS-GKQTAQNAIWIDCGIH 67
Db 121 YHTLEISQEMDLVAEHPLGLSVKNIGSSFENRPMVLKFTSGGD---KPAIWLDAIGH 177
QY 68 AREWISPAFCWIFG-----HNRWRKN 90
Db 178 AREWVQTALWTANKIVSYGKDPSITSLDALDIFLLVTPNPDGVFSQTKNRMRKT 237
QY 91 RSYVANNHCITGLNRFNASKWCEGASSSCSETYCGLYPSEPEVKAVASFLARNIN 150
Db 238 RSKVSGSLCVGVDPRNW-DAGGGGAGSNPCSDSYHGSANSEVEKSVIVDFIKSH-G 295
QY 151 QIKAYISMHSYSHIVFPYYSYRSKSDHEELSLVASEAVRAIEKTSKNTRYTHGSET 210
Db 296 KVKAFITLHSYSQLLMFPYGYKCTKLDLDFDELSEVAQAQSL-RSLHGTKYKVGPICSV 354
QY 211 LYLPAGGGDDWIYDLGKISFTSN-----PPVEKLLP 242
Db 355 IYQAGSGSIDWSYDYGKYSFAFELRDTGRYGLFLPARQILP 396

RESULT 14
CBP4_HUMAN STANDARD; PRT; 421 AA.
AC Q9UI42;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN CPA3 OR CPA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99310537; PubMed=10383164;
RX Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT deacetylase inhibitors during differentiation of prostate epithelial
RT cancer cells.";
RL Cancer Res. 59:2981-2988(1999).
RN [2]
RP IMPRINTING.
RX MEDLINE=22483629; PubMed=12552318;
RA Kayashima T., Yamasaki K., Yamada T., Sakai H., Miwa N., Ohta T.,
RA Yoshitake K.-I., Matsumoto N., Nakane Y., Kanetake H., Ishino F.,
RA Niihara N., Kishino T.;
RT "The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32
RT imprinting domain.";
RL Hum. Genet. 112:220-226(2003).
CC -1- FUNCTION: Could be involved in the histone hyperacetylation
CC pathway.
CC -1- TISSUE SPECIFICITY: Fetal expression in the adrenal gland, brain,
CC heart, intestine, kidney, liver and lung. Except for fetal brain
CC that shows no imprinting, expression was found preferentially from
CC the maternal allele.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF095719; AAF23230.1; -.
CC HSSP; P48052; IAYE.

DR MEROPS; M14.017; -.
DR Genew; HGNC:15740; CPA4.
DR MIM; 607635; -.
DR GO; GO:0004182; F:carboxypeptidase A activity; NAS.
DR GO; GO:0016573; P:histone acetylation; NAS.
DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn.carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn.carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT PROPEP 17 113 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 114 421 CARBOXYPEPTIDASE A4.
FT METAL 181 181 ZINC (BY SIMILARITY).
FT METAL 184 184 ZINC (BY SIMILARITY).
FT METAL 308 308 ZINC (BY SIMILARITY).
FT ACT_SITE 382 382 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 250 273 BY SIMILARITY.
SQ SEQUENCE 421 AA; 47379 MW; 91883D25B705BC6A CRC64;

Query Match 34.2%; Score 458; DB 1; Length 421;
Best Local Similarity 38.4%; Pred. No. 5.2e-35;
Matches 103; Conservative 35; Mismatches 82; Indels 48; Gaps 6;

QY 6 YEQYHSLNLIYSWIEFITERHPDMLTKIHIGSFKEYPLVVLKVSKEQTAKNAIWIDCG 65
Db 120 YGAYHSLNLIYSWIEFITERHPDMLTKIHIGSFKEYPLVVLKVSKEQTAKNAIWIDCG 179
QY 66 IHAREWISPAFCW-----FI-----GHNRMWR 88
Db 180 IHSREWISQATATWTARKIVSDYQDRPAITSILEKMDIFLLPVANPDGVYVYTOTQNLWR 239
QY 89 KNRSFYANNHCITGLNRFNASKWCEGASSSCSETYCGLYPSEPEVKAVASF 144
Db 240 KTRSRNPGSCIGADPNRNWNASFAGK-----GASDNPCEVYVGHGPHANSEVEKSVDF 294
QY 145 LRNNIQAISYMSHSYSHIVFPYYSYRSKSDHEELSLVASEAVRAIEKTSKNTRYTH 204
Db 295 IQKHGN-FKGFDLHSYSQLLMFPYGYKSVKVKAPDAEELQKVARLAALAKALASVS-GTEYQV 352
QY 205 GHGSETLYLAPGGDDWIYDLGKISYFT 232
Db 353 GPTCTTVYPASGSSIDWAYDNGIKFAFT 380

RESULT 15
CBPZ_SIMVI STANDARD; PRT; 304 AA.
AC P42788;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC Simuliidae; Simulium.
OX NCBI_TaxID=7192;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gut;
RX MEDLINE=94093864; PubMed=8269093;
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RL trypsin-like and carboxypeptidase-like proteins.";
RL Insect Mol. Biol. 1:149-163(1993).
CC -1- FUNCTION: Involved in the digestion of the blood meal.
CC -1- SUBCELLULAR LOCATION: Secreted.
```

```
CC -!- TISSUE SPECIFICITY: Gut-specific.
CC -!- INDUCTION: By blood meal.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08481; AAA18531.1; -
DR HSSP; P48052; LAYE.
DR MEROPS; M14.0PA; -.
DR InterPro; IPR000834; Zn_carboOpept.
DR Pfam; PF00246; Zn_carboOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT NON_TER 1
FT METAL 58 58 ZINC (BY SIMILARITY).
FT METAL 61 61 ZINC (BY SIMILARITY).
FT METAL 184 184 ZINC (BY SIMILARITY).
FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 125 148 BY SIMILARITY.
SQ SEQUENCE 304 AA; 34849 MW; 2E6B3FF8A6AA9144 CRC64;

Query Match 30.8%; Score 411.5; DB 1; Length 304;
Best Local Similarity 33.3%; Pred. No. 7.2e-31;
Matches 94; Conservative 52; Mismatches 85; Indels 51; Gaps 9;

QY 8 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFKEKYPPLYVLKVGKEQTAKNAIWIIDCGIH 67
Db 1 QYHTLPEIYSWLDRLVQEHVEHPVVGKSYEGREIRGVKSYKK--GNPVVWVESNIH 58

QY 68 AREWISPA---FCL-----WFI-----GH-----NRMWRKNRS 92
Db 59 AREWITAATTYLLNELLTSKNTIREMAENYDWIYPPVTNPDGYVYVTHTTDRMRKTRS 118

QY 93 FYANNHCIGTDLNRNFASKHWCBEGASSSCSETYCGLYPESEPEVKAVASFLRRNIQI 152
Db 119 PNPDSLCAGTDPNRNW-NPHMWEQGTSSRCTETGYGKKAFSEVETRSFDFLTKGQI 177

QY 153 KAYISMHSYSQHLVFPYSYTRSKSDHELSLVAESAIVRAIEKTSKNTRYTHGHGETLY 212
Db 178 KVLAFHSYSQQLLPFYGYHTCQHTYHNDLQALGDAARSLAQ-RYGTDTYVGNIDAIY 236

QY 213 LAPGGDDWTY---DLGIKYSFTSNP-----PVEKLLP 242
Db 237 PASGGMDWAYDTLDIPIATYIELRPDGNWGNFQLPANQIIP 278
```

Search completed: January 1, 2004, 19:29:01
Job time : 10.7578 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:19:56 ; Search time 41.5992 Seconds
(without alignments)
1526.015 Million cell updates/sec

Title: US-09-980-881A-4

Perfect score: 1338

Sequence: 1 ASASYEQYHSLNEIYSWIE.....IKYFTSNPPVVKLPLSLK 246

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mnc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1327	99.2	360	Q9P2Y6	Q9p2y6 homo sapien
2	1239.5	92.6	423	Q15114	Q15114 homo sapien
3	1234.5	92.3	423	Q961Y4	Q961y4 homo sapien
4	1105.5	82.6	422	Q9JHH6	Q9jhh6 mus musculus
5	1095.5	81.9	422	Q9EGV9	Q9egv9 rattus norv
6	1095.5	81.9	422	Q9QZFO	Q9qzf0 mus musculus
7	622.5	46.5	408	Q8AXN3	Q8axn3 paralichthy
8	612.5	45.8	416	Q9RUF2	Q9rfu2 bothrops ja
9	582	43.5	416	Q9XSP3	Q9xsp3 sus scrofa
10	557	41.6	412	Q9P7597	Q9p7597 rattus norv
11	550.5	41.1	374	Q8IVL8	Q8ivl8 homo sapien
12	547	40.9	437	Q8N4T0	Q8n4t0 homo sapien
13	544.5	40.7	434	Q8AVS2	Q8avs2 xenopus lae
14	514.5	38.5	419	Q8UUK1	Q8uuk1 gallus gall
15	488.5	36.5	419	Q9TV85	Q9tv85 sus scrofa
16	486.5	36.4	436	Q8R4H4	Q8r4h4 mus musculus

17	484.5	36.2	262	11	Q8BVD0	Q8bvd0 mus musculus
18	477.5	35.7	247	4	Q96KZ9	Q96kz9 homo sapien
19	476	35.6	420	11	Q8BMK6	Q8bmk6 mus musculus
20	465.5	34.8	436	4	Q8WXQ8	Q8wxq8 homo sapien
21	456.5	34.1	419	13	Q8AXN4	Q8axn4 paralichthy
22	456	34.1	198	4	Q9NTI8	Q9nti8 homo sapien
23	449.5	33.6	226	4	Q8TDE8	Q8td8 homo sapien
24	428	32.0	452	13	Q8UW45	Q8uw45 fugu rubrip
25	426	31.8	1132	5	Q9W475	Q9w475 drosophila
26	424	31.7	424	5	Q9VL87	Q9vl87 drosophila
27	424	31.7	428	5	Q8IGC4	Q8igc4 drosophila
28	419	31.3	351	5	Q8TQ01	Q8tq01 drosophila
29	419	31.3	422	5	Q9VLZ2	Q9vlz2 drosophila
30	419	31.3	445	5	Q9W478	Q9w478 drosophila
31	406.5	30.4	419	13	Q8AXN5	Q8axn5 paralichthy
32	403	30.1	323	5	Q19121	Q19121 caenorhabdi
33	402	30.0	416	5	Q9VRZ3	Q9vrz3 drosophila
34	396.5	29.6	584	5	Q9TZH1	Q9tzh1 caenorhabdi
35	385.5	28.8	453	5	Q9VJ42	Q9vj42 drosophila
36	385.5	28.8	454	5	Q9TZC6	Q9tzc6 caenorhabdi
37	382.5	28.6	341	5	Q8SXR8	Q8sxr8 drosophila
38	382.5	28.6	424	5	Q9VS66	Q9vs66 drosophila
39	380	28.4	430	5	Q9VL86	Q9vl86 drosophila
40	379	28.3	351	4	Q8NA08	Q8na08 homo sapien
41	377.5	28.2	423	5	Q61532	Q61532 drosophila
42	375	28.0	427	5	Q9U9K2	Q9u9k2 aedes aegypt
43	374.5	28.0	488	5	Q9BL88	Q9bl88 caenorhabdi
44	373	27.9	430	5	Q8SZP6	Q8szp6 drosophila
45	370.5	27.7	419	5	Q9VLZ1	Q9vlz1 drosophila

ALIGNMENTS

RESULT 1

Q9P2Y6	PRELIMINARY;	PRT;	360 AA.
AC	Q9P2Y6;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-WAR-2003 (TREMBLrel. 23, Last annotation update)		
DE	Carboxypeptidase B-like protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Matsumoto A.;		
RT	"Isolation, molecular cloning, and partial characterization of a novel		
RT	carboxypeptidase B from human plasma.";		
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92042093; PubMed=1939207;		
RA	Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;		
RT	"Isolation, molecular cloning, and partial characterization of a novel		
RT	carboxypeptidase B from human plasma.";		
RL	J. Biol. Chem. 266:21833-21838(1991).		
DR	EMBL; AB011969; BAA90475.1; -		
DR	HSSP; P00730; 2CTC.		
DR	MEROPS; M14.009; -		
DR	InterPro; IPR003146; Propep M14.		
DR	InterPro; IPR000834; Zn_CarboPept.		
DR	Pfam; PF02244; Propep_M14; 1.		
DR	Pfam; PF00246; Zn_CarboPept; 1.		
DR	PRINTS; PR00765; CRBOXYPTASEA.		
DR	SMART; SM00631; Zn_pept; 1.		
SQ	SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;		

Query Match 99.2%; Score 1327; DB 4; Length 360;
Best Local Similarity 99.2%; Pred. NO. 1e-119;
Matches 244; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSQKEQTAKNAI	60
Db	115	ASASYEQVHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSQKEQAANKAI	174
Qy	61	WDTCGHAREWISPAFLCWFIGHNRMRKNRSFYANNHCITGDLNRFASKHWCBEASS	120
Db	175	WDTCGHAREWISPAFLCWFIGHNRMRKNRSFYANNHCITGDLNRFASKHWCBEASS	234
Qy	121	SSCSETYCGLYPSEPEVKAVASFLRRNINOIKAYISMHSYSQHIVFPYSYTRSKSKDHE	180
Db	235	SSCSETYCGLYPSEPEVKAVASFLRRNINOIKAYISMHSYSQHIVFPYSYTRSKSKDHE	294
Qy	181	EUSLVASAARAIEKTSKNTRYTHCHGSETUYLAPGGDDDWIYDLGIKYSFTSNPPVEKL	240
Db	295	EUSLVASAARAIEKISKNTRYTHCHGSETUYLAPGGDDDWIYDLGIKYSFTSNPPVEKL	354
Qy	241	LPLSLK 246 	
Db	355	LPLSLK 360 	

RESULT 2

Q15114		PRELIMINARY;	PRT;	423 AA.
ID	Q15114			
AC	Q15114;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	PCPB protein.			
DE	PCPB.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata;			
OC	Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCB_I_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RA	MEDLINE=92042093; PubMed=1939207;			
RX	Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;			
RT	"Isolation, molecular cloning, and partial characterization of a novel			
RL	carboxypeptidase B from human plasma.";			
J. Biol. Chem. 266:21833-21838(1991).				
EMBL; M75106; AAA60042.1; .				
DR	HSP; P00730; 1PYT.			
DR	MEROPS; M14_009; .			
DR	Newgen; HGNC:2300; CPB2.			
DR	InterPro; IPR003146; Propep M14.			
DR	InterPro; IPR000834; Zn_carboOsept.			
DR	Pfam; PF02244; Propep M14; 1.			
DR	Pfam; PF00246; Zn_carboOsept; 1.			
DR	PRINTS; PR00745; CRBOXYPIASEA.			
DR	SMART; SM00631; Zn_sept; 1.			
DR	CHAIN 23 423			
SF	SEQUENCE 423 AA; 48442 MW; 9B383272F6EE79F4 CRC64;			
FT	PLASMA CARBOXYPEPTIDASE B.			

Query Match	92.6%	Score 1239.5;	DB 4;	Length 423;
Best Local Similarity	86.2%;	Pred. No. 3.6e-111;		
Matches 235;	Conservative 0;	Mismatches 0;	Indels 37;	Gaps 1;
QY	1	ASASYEQVHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPILYLKVSGKEQTAKNAI	60	
Db	115	ASASYEQVHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPILYLKVSGKSEQTAKNAI	174	
QY	61	WIDCGIHAREWIISPAFCFLWFIGH-----	83	
Db	175	WIDCGIHAREWIISPAFCFLWFIGHTPYGIIGQYTNLRLRVDFVMFPVVNDGYDIYSWK	234	
QY	84	NRMWRKNRSPYANNHCIGTDLNRNFASKHWCEGASSSCSSETCYGLYPESEPEVKAVAS	143	
Db	235	NRMWRKNRSPYANNHCIGTDLNRNFASKHWCEGASSSCSSETCYGLYPESEPEVKAVAS	294	
QY	144	FLRRNINOKAYITSMHSYSOHIHVFPVSATRSKSKDHHEELSLIVASEAVRAJEKTSKONTRYT	203	

Db	295	FLRNNINQIKAYISMHSYSQHVPPYSYTRSKSKDHELSVLVAEAVRAIKETSKNTRYT	354
Qy	204	HGHGSETLYLAPGGGDDWIYDLGIKYSFT	232
Db	355	HGHGSETLYLAPGGGDDWIYDLGIKYSFT	383
RESULT 3			
ID	Q961Y4	PRELIMINARY;	PRT; 423 AA.
AC	Q961Y4;		
DT	01-DEC-2001	(TrEMBLrel. 19, Created)	
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Carboxypeptidase B2 (Plasma).		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
NCBI_Taxid=	9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

RESULT 4

Q9JHH6		PRELIMINARY;	PRT;	422 AA.
ID	Q9JHH6			
AC	Q9JHH6;			
DT	01-OCT-2000	(TREMBLrel. 15, Created)		
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)		
DT	01-WAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)			
DE	(I110032P04RIK protein).			
CS	CPB2 OR TAFI OR I110032P04RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Eumalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20201996; PubMed=10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
van Rossum A.G.S.H., Gebbink M.F.G.B., Weijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
inhibitor.";
RL Thromb. Haemost. 83:297-303 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
Schröter L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
Yamashiro-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR ENBL; AB021968; BAB03402.1; -;
DR ENBL; AF164524; AAF62385.1; -;
DR ENBL; AK004045; BAB23141.1; -;
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -;
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

Query Match 82.6%; Score 1105.5; DB 11; Length 422;
Best Local Similarity 77.0%; Pred. No. 3.1e-98;
Matches 207; Conservative 10; Mismatches 15; Indels 37; Gaps 1;
1 ASASYEQYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAKNAI 60
114 ASASYEQYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAKNAI 173
61 WIDCGIHAREWISPAFLWFIGH----- 83
174 WIDCGIHAREWISPAFLWFIGHYTFQHGKENTYTRLLRHVDYIMPVNMVDGYDTWK 233
84 NRWRKNRSFYANNHCIGTDLNRNFASKHWCSEGASSSCSCTCYGLYPESEPEVKAVAS 143
234 NRWRKNRSAHKNRCVGTDLNRNFASKHWCSEGASSSCSCTCYGLYPESEPEVKAVAD 293
144 FLRRNIQIKAYISMHSYQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203

Db 294 FLRRNIQIKAYISMHSYQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 353
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
Db 354 HGHGSETLYLAPGGDDWIYDLGIKYSFT 382
RESULT 5
Q9EQV9 PRELIMINARY; PRT; 422 AA.
ID AC Q9EQV9;
DC 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
OS Pre-procarboxypeptidase R.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20471387; PubMed=11021404;
RA Kato T., Sato T., Mateuo S., Yamamoto T., Campbell W., Hotta N.,
Okada N., Okada H.;
RT "Molecular cloning and partial characterization of rat
procarboxypeptidase R and carboxypeptidase N.";
RL Microbiol. Immunol. 44:719-728 (2000).
DR ENBL; AB042598; BAB18617.1; -;
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -;
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR SMART; SM00631; Zn_pept; 1.
KW Carboxypeptidase.
SQ SEQUENCE 422 AA; 48826 MW; PFFD32A51A9366C8 CRC64;

Query Match 81.9%; Score 1095.5; DB 11; Length 422;
Best Local Similarity 75.8%; Pred. No. 2.8e-97;
Matches 204; Conservative 11; Mismatches 17; Indels 37; Gaps 1;
1 ASASYEQYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAKNAI 60
114 ASASYEQYHSLNEIYSWIEFIERHPDMLTKIHGSSFEKYPYLVKVGSGEQTAKNAI 173
61 WIDCGIHAREWISPAFLWFIGH----- 83
174 WIDCGIHAREWISPAFLWFIGHYTFQHGKENTYTRLLRHVDYIMPVNMVDGYDTWK 233
84 NRWRKNRSFYANNHCIGTDLNRNFASKHWCSEGASSSCSCTCYGLYPESEPEVKAVAS 143
234 NRWRKNRSVHMNRVCVGTDLNRNFASKHWCSEGASSSCSCTCYGLYPESEPEVKAVAD 293
144 FLRRNIQIKAYISMHSYQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
294 FLRRNIQIKAYISMHSYQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 353
QY 204 HGHGSETLYLAPGGDDWIYDLGIKYSFT 232
Db 354 HGHGSETLYLAPGGDDWIYDLGIKYSFT 382
RESULT 6
Q9QZF0 PRELIMINARY; PRT; 422 AA.
ID AC Q9QZF0;
DC 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
OS Carboxypeptidase U.
GN Cpb2.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J;
RA He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RT gene";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF186188; AAF00528.1; -
DR HSSP; P00730; 5CPA.
DR MEROPS; M14.009; -
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep M14.
DR IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_dept; 1.
SQ SEQUENCE 422 AA; 48832 MW; B53FFB09943954E CRC64;

Query Match 81.9%; Score 1095.5; DB 11; Length 422;
Best Local Similarity 76.2%; Pred. No. 2.8e-97;
Matches 205; Conservative 11; Mismatches 16; Indels 37; Gaps 1;

QY 1 ASASYEQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYVLKVGKEQTAKNAI 60
DB 114 ASASYEQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYVLKVGKEQTAKNAI 173
QY 61 WIDCGIHAREWISPAFLWFIH-----83
DB 174 WIDCGIHAREWISPAFLWFIHVGTVQFHGKENVYTRLLRHVDYIMPVNVGDYDTWK 233
QY 84 NRMWRKNSFYANNHCIGTDLNRFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVAS 143
DB 234 NRMWRKNSAHKNRCVGTDLNRFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVAD 293
QY 144 FLRRNIQIKAYISMHSYSHQIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203
DB 294 FLRRNIQIKAYISMHSYSHQIVFPYSTRSKSDHEELSLVASEAVRAIGSFNKNTRYT 353
QY 204 HGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 354 HGSGSESLLAPGGDDWIYDLGIKYSFT 382

RESULT 7
Q8AXN3 PRELIMINARY; PRT; 408 AA.
AC Q8AXN3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B (Fragment).
GN CPB.
OS Paralicthys olivaceus (flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Pleuronectiformes;
OC Pleuronectoidae; Paralicthyidae; Paralicthys.
OX NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RA Srivastava A.S., Kurokawa T., Suzuki T.;
RT "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT A2 and B from the Japanese flounder, Paralicthys olivaceus";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB099302; BAC53789.1; -
FT NON_TER 1
SQ SEQUENCE 408 AA; 45904 MW; 809CAACD9D2DA34B CRC64;

Query Match 46.5%; Score 622.5; DB 13; Length 408;

Best Local Similarity 45.8%; Score 612.5; DB 13; Length 416;
Matches 121; Conservative 44; Mismatches 63; Indels 41; Gaps 5;

QY 2 SASY-YEQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYVLKVGKEQTAKNAI 60
DB 111 TAGYNEKYNWEKIDAWTADIANENPSLVSRLOIGTTTPEGPMPLKV-GKPGVKNKAI 169
QY 61 WIDCGIHAREWISPAFLWFI-----GH 83
DB 170 FIDCGPHAREWISPAFLWFIHVGTVQFHGKENVYTRLLRHVDYIMPVNVGDYDTWK 229
QY 84 NRMWRKNSFYANNHCIGTDLNRFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVAS 143
DB 230 SRMWRKNSVAGSTCIGTDLNRFASKHWCCEGASSSSCSETYCGSKPESEKETALAD 288
QY 144 FLRRNIQIKAYISMHSYSHQIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203

Best Local Similarity 46.1%; Pred. No. 1.2e-51;
Matches 123; Conservative 41; Mismatches 64; Indels 39; Gaps 4;

QY 3 ASYEQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYVLKVGKEQTAKNAIWI 62
DB 103 AHSYTKYNNLDKQVSWIASISSNPDLISKQVIGNTYGRPMPTVLKLGKSSSTKPAIFM 162
QY 63 DCGIHAREWISPAFLWFI-----GH-----NR 85
DB 163 DCGIHAREWISPAFLWFIHVGTVQFHGKENVYTRLLRHVDYIMPVNVGDYDTWK 222
QY 86 NRMWRKNSFYANNHCIGTDLNRFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVASFL 145
DB 223 NRMWRKNSRKSQSGCIGADPNRF-DAGNCSMGASDNPCSDTFCGYTPESEIEVKVADFI 281
QY 146 RRNIQIKAYISMHSYSHQIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYTHG 205
DB 282 RRNRSIKAYLTVHSYQLLFFYSYQLAADHSELKMKVAGASAL-RSLYGTYYTSG 340
QY 206 HGSETLYLAPGGDDWIYDLGIKYSFT 232
DB 341 PGATTIYPAAGGDDWAYDLGVKYSY 367

RESULT 8
Q9PUP2 PRELIMINARY; PRT; 416 AA.
AC Q9PUP2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase homolog.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT "Screening of Bothrops jararaca pancreas cDNA library";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190274; AAF01344.1; -
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.003; -
DR InterPro; IPR003146; Propep M14.
DR IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_dept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 416 AA; 47723 MW; 4F999854DD72B7A7 CRC64;

Query Match 45.8%; Score 612.5; DB 13; Length 416;
Best Local Similarity 45.0%; Pred. No. 1.2e-50;
Matches 121; Conservative 44; Mismatches 63; Indels 41; Gaps 5;

QY 2 SASY-YEQHSLNEIYSWIEFITERHPDMLTKIHGSSFEKPYLYVLKVGKEQTAKNAI 60
DB 111 TAGYNEKYNWEKIDAWTADIANENPSLVSRLOIGTTTPEGPMPLKV-GKPGVKNKAI 169
QY 61 WIDCGIHAREWISPAFLWFI-----GH 83
DB 170 FIDCGPHAREWISPAFLWFIHVGTVQFHGKENVYTRLLRHVDYIMPVNVGDYDTWK 229
QY 84 NRMWRKNSFYANNHCIGTDLNRFASKHWCCEGASSSSCSETYCGLYPESEPEVKAVAS 143
DB 230 SRMWRKNSVAGSTCIGTDLNRFASKHWCCEGASSSSCSETYCGSKPESEKETALAD 288
QY 144 FLRRNIQIKAYISMHSYSHQIVFPYSTRSKSDHEELSLVASEAVRAIEKTSKNTRYT 203

```
Db 289 FIRNRSLIQLVTHSYQMLLPYSYTYDLTSSNNKLNLSIAKEAIREL-KVLFGTEYT 347
QY 204 HGHGSETLYLAPGGDDWYDLGIKYSFT 232
Db 348 YGFGAATYPAAGGDDWAYDOGIKRAFT 376

RESULT 9
Q9XSP3
ID Q9XSP3 PRELIMINARY; PRT; 416 AA.
AC Q9XSP3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Procarboxypeptidase B precursor (EC 3.4.17.2).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=99312926; PubMed=10391940;
RA Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA Hersberger C., Aviles F.;
RT "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL J. Biol. Chem. 274:19925-19933 (1999).
DR EMBL: AJ133775; CAB46991.1; -.
DR HSSP: P09955; 1PBA.
DR InterPro: IPR003146; Propep M14.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF02244; Propep M14; 1.
DR Pfam: PF00246; Zn_carbopept; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR SMART: SM00631; Zn_pept; 1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Hydrolase; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 416 PROCARBOXYPEPTIDASE B.
SQ SEQUENCE 416 AA; 47381 MW; 5FA7CED7B45A6A CRC64;

Query Match 43.5%; Score 582; DB 6; Length 416;
Best Local Similarity 44.3%; Pred. No. 1e-47;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIFITERHPDMLTKIHGSSFEKYPYLVKVGSGKEQTAKNAIWDCG 65
Db 116 YEKYNWETIEATKQVTSNPDLISRTAIGTFLGNIIYLLKV-GKPGPNKPAIFMDCG 174
QY 66 IHAREWISPAFLWFI-----GH-----NRMR 88
Db 175 FHAREWISPAFCQFVREAVLTYGSHMTFELNKLDFYVPLVNLIDGIYITKRMWR 234
QY 89 KNRSPYANNHCIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 148
Db 235 KTRSTNAGTTCIGTDPRNF-DAGWCCTGTASTDPCDTEYCGSAAESEKETKALADFIRN 293
QY 149 INQIKAYISMHSYSHIVFPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db 294 LSSIKAYLTTHSYQMLLPYSYDYKLPENNAELNLAKAAVKEL-ATLYGTYKTYGPGA 352
QY 209 ETLYLAPGGDDWYDLGIKYSFT 232
Db 353 TTYTPAAGGDDWAYDOGIKYSFT 376

RESULT 10
P97597
ID P97597 PRELIMINARY; PRT; 412 AA.
AC P97597;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
```

```
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mast cell carboxypeptidase A precursor (Fragment).
GN R-CPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RX MEDLINE=97149430; PubMed=8996238;
RA Lutzelschwab C., Rejler G., Aveskogh M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations.";
RL J. Exp. Med. 185:13-29 (1997).
DR EMBL: U67914; AAB48267.1; -.
DR HSSP: P09955; 1NSA.
DR InterPro: IPR003146; Propep M14.
DR InterPro: IPR000834; Zn_carbopept.
DR Pfam: PF02244; Propep M14; 1.
DR Pfam: PF00246; Zn_carbopept; 1.
DR PRINTS: PR00765; CRBOXYPTASEA.
DR SMART: SM00631; Zn_pept; 1.
DR PROSITE: PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Signal.
FT NON TER 1 1
FT SIGNAL <1 10 POTENTIAL.
FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
SQ SEQUENCE 412 AA; 47944 MW; DAB5955FC49137D CRC64;

Query Match 41.6%; Score 557; DB 11; Length 412;
Best Local Similarity 43.0%; Pred. No. 2.6e-45;
Matches 113; Conservative 38; Mismatches 72; Indels 40; Gaps 5;

QY 6 YEQYHSLNEIYSWIFITERHPDMLTKIHGSSFEKYPYLVKVGSGKEQTAKNAIWDCG 65
Db 111 YAKYDNWKNIVSWTERKVEKHPENSRIGKISTVEDNPPLYVKI-GRKDGKRAIFMDCG 169
QY 66 IHAREWISPAFLWFI-----GH-----NRMR 88
Db 170 IHAREWISPAFCQFVYQAAKSYGKNKIMTKLDRNFVYVFNVDGYIWSWTKDRMR 229
QY 89 KNRSPYANNHCIGTDLNRNPFASKHWCCEGASSSCSETYCGLYPESEPEVKAVASFLRN 148
Db 230 KNRKPNSTCIGTDLNRNF-DVSWDSSPNTPCLSVYRGPAPESEKETKAVTNFIRSH 288
QY 149 INQIKAYISMHSYSHIVFPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGHS 208
Db 289 LNSIKAYITPHSYQMLLPYGYTIKLPNHNQDLKVIATDVL-SSRYETRYIYGPIA 347
QY 209 ETLYLAPGGDDWYDLGIKYSF 231
Db 348 STIYKTSGLDWAYDLGIKHTF 370

RESULT 11
Q8IVL8
ID Q8IVL8 PRELIMINARY; PRT; 374 AA.
AC Q8IVL8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Zn-carboxypeptidase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
```


RP SEQUENCE FROM N.A.
RC STRAIN-domesticus; TISSUE=Pancreas;
RA Hasegawa S., Honda K., Hikami Y.;
RT "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A
precursor";
RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL: X64539; CAA45837.1; -
DR MEROPS; M14.001; -
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Hydrolase.
SQ SEQUENCE 419 AA; 47235 MW; 84B4CB557B714FC1 CRC64;
Query Match 36.5%; Score 514.5; DB 13; Length 419;
Best Local Similarity 42.0%; Pred. No. 3.4e-41;
Matches 110; Conservative 40; Mismatches 73; Indels 39; Gaps 5;
QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPLVYLKVSKEQTAKN--AIWI 65
DB 121 YTSYHTLEETIYAFMDLVAENPNLVSKEIGRTTENRPIYVLKFS-KGGTNRPAIWDTG 179
QY 66 IHAREWISPAFLWFI-----GHNRRWKN 90
DB 180 IHSREWVTQASGVWFVAKKIVQEDGLANLDMQIDFLEIVTNPDPGFATHTSNRWRKT 239
QY 91 RSPYANNHCITGLNRFNFKWCEBEGASSSCSETCYGLYPESEPEVKAVASFLARNIN 150
DB 240 RSKRSGSLCVGVDPNRW-DAGFGGAGASSNPCSETHYGFYANSEPEVKAVIDFVKSHKN 298
QY 151 QIKAYISMSHSYQHIVFPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHGSET 210
DB 299 -IKAFISHSYQLLYPYGYTTAVPDKEELQVAKAVAL--SSLYGTNYKYGSIIIT 356
QY 211 LYLAPGGDDMIYDLGIKYSFT 232
DB 357 IYQASGGTIDWTYNOGIKYSFT 378
RESULT 15
Q9TV85 PRELIMINARY; PRT; 419 AA.
AC Q9TV85;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase A1 (EC 3.4.17.1).
GN PCPAl.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99192816; PubMed=10092856;
RA Darnie S., Juge N., Marino C., Aviles F.X., Puigserver A., Chaix J.C.,
RA Guo X.J.;
RT "Cloning, sequencing and functional expression of a cDNA encoding
porcine pancreatic preprocarboxypeptidase A1.";
RL Eur. J. Biochem. 259:719-725(1999).
DR EMBL; AF076222; AAD17690.1; -
DR HSSP; P09954; lPCA.
DR InterPro; IPR003146; Propep M14.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF02244; Propep M14; 1.
DR Pfam; PF00246; Zn_carbopept; 1.

DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Hydrolase.
SQ SEQUENCE 419 AA; 47235 MW; 84B4CB557B714FC1 CRC64;
Query Match 36.5%; Score 488.5; DB 6; Length 419;
Best Local Similarity 39.0%; Pred. No. 1.1e-38;
Matches 104; Conservative 42; Mismatches 74; Indels 47; Gaps 6;
QY 6 YEQYHSLNEIYSWIEFITERHPDMLTKIHGSSFEKYPLVYLKVSKEQTAKN--AIWI 62
DB 119 YATYHTLEETIYDFMDILVAEHPOLVSKLQIGSSYEGRPYIYVLKFS---TGGNNRPAIWI 174
QY 63 DCGIHAREWISPAFLWFI-----GHNRR 85
DB 175 DTGHSREWVTQASGVWFVAKKITEDYQDPAFATLNDLIDFLEIVTNPDPGFATHTSEN 234
QY 86 MWRKNRSFYANNHCITGLNRFNFKWCEBEGASSSCSETCYGLYPESEPEVKAVASFL 145
DB 235 MWRKTRSRSTSGFCVGVDPNPNW-DAGFGGAGASSNPCSETHYHGKFPNSEVEVKSIYDFV 293
QY 146 RRNINOIKAYISMSHSYQHIVFPYSYTRSKSDHELSLVASEAVRAIEKTSKNTRYTHG 205
DB 294 NDHGN-IKAFISHSYQLLYPYGYTAPADKDELQISKSAVAL-TSLYGTQPKYQ 351
QY 206 HGSETLYLAPGGDDMIYDLGIKYSFT 232
DB 352 SIITTIYQASGGTIDWTYNOGIKYSFS 378
Search completed: January 1, 2004, 19:31:58
Job time : 43.5992 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:34:47 ; Search time 3798.88 Seconds
(without alignments)
2649.143 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYEQYHLSNLEIYSWIE.....IKYSTSNPPVEXKLPLSLK 246

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-O=/cpn2 1/USPTO spool/US09980881/runat_31122003_135915_26829/app_query.fasta_1.1628
-DB=GenEmbl -OPWT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09980881 @CCN 1 1 8325 @runat_31122003_135915_26829 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPHLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.Or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.ste.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1327	99.2	1546	9	AB011969	Homo sapi
2	1289.5	96.4	1625	6	AR232183	Sequence
3	1272	95.1	1272	6	AR086324	Sequence
4	1272	95.1	1272	6	BD106653	Method of
5	1272	95.1	1728	6	AX409750	Sequence
6	1272	95.1	1728	6	AX706330	Sequence
7	1272	95.1	1728	9	HUMPCPBX	Human prepr
8	1272	95.1	1749	6	I16100	Sequence 2
9	1272	95.1	1749	6	I33526	Sequence 2
10	1267	94.7	1272	9	BT006936	Homo sapi
11	1267	94.7	1272	12	BT007990	Synthetic
12	1267	94.7	1715	9	BC007057	Homo sapi
13	1112	83.1	1421	10	AB021968	Mus muscu
14	1112	83.1	1490	10	AF164524	Mus muscu
15	1102	82.4	1430	10	AF186188	Mus muscu
16	1099	82.1	1425	10	AB042598	Rattus no
17	622.5	46.5	1441	5	AB099302	Paralicht
18	612.5	45.8	1327	5	AF190274	Bothrops
19	610	45.6	927	6	AR072911	Sequence
20	582	43.5	921	6	I67699	Sequence 3
21	582	43.5	1215	6	I67698	Sequence 1
22	582	43.5	1251	4	SSC133775	Sus scrof
23	578	43.2	1251	4	DOGZAP47	Canis famil
24	567	42.4	999	6	A51908	Sequence 72
25	567	42.4	999	6	AR085871	Sequence
26	567	42.4	1053	6	A51904	Sequence 68
27	567	42.4	1053	6	AR085868	Sequence
28	567	42.4	1254	12	BT008111	Synthetic
29	567	42.4	1263	6	A51896	Sequence 60
30	567	42.4	1263	6	A87530	Sequence 11
31	567	42.4	1263	6	AR085861	Sequence
32	567	42.4	1284	6	A51913	Sequence 77
33	567	42.4	1284	6	AR085875	Sequence
34	567	42.4	1310	9	HS224866	Homo sapi
35	567	42.4	1454	9	BC015338	Homo sapi
36	561	41.9	1442	10	MUSCARA	Mouse mast
37	560	41.9	1053	6	AR300428	Sequence
C 38	560	41.9	1053	6	AR300429	Sequence
39	560	41.9	1053	6	AX000417	Sequence
C 40	560	41.9	1053	6	AX000418	Sequence
41	560	41.9	1053	6	BD080205	Expressio
C 42	560	41.9	1053	6	BD080206	Expressio
43	560	41.9	1059	6	A51917	Sequence 81
44	560	41.9	1059	6	AR085878	Sequence
45	559	41.8	1059	6	A51921	Sequence 85

ALIGNMENTS

RESULT 1

```

AB011969      1546 bp      mRNA      linear      PRI 02-FEB-2000
LOCUS      Homo sapiens mRNA for carboxypeptidase B-like protein, complete
DEFINITION
ACCESSION      AB011969
VERSION      AB011969.1 GI:6855463
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE      92042093
PUBMED      1939207
REFERENCE
2 (sites)
Matsumoto,A.
Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
Unpublished
REFERENCE
3 (bases 1 to 1546)
Matsumoto,A.
Direct Submission
Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
Medicine, Department of Radiation Biophysics & Genetics;
Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
(E-mail:amat@med.kobe-u.ac.jp, Tel:81-78341-7451)
Sequence updated (10-Apr-1998).
COMMENT
Location/Qualifiers
1..1546
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
20..1102
/notes="carboxypeptidase B-like protein in human brain"
/codon_start=1
/product="carboxypeptidase B-like protein"
/protein_id="BAA90475.1"
/db_xref="GI:6855464"
/translations="MKLSLAVLPVILFCEQHFVAFQTVGQVLAALPRTSROVOVLON
LTWTVELVWQPVTDLIKKQVHFVNADVNDVKHLNYSIGIPCSVLADVEDLI
QQQISNDTSPRASAYEYQYHSLNIEYSEIFETTERPDMLTKIHIGSFEKPYLYV
LKVSQEQAANAIDCGIHAREWISPAFLWFIQHNRRWRKRSFYANNHCIGFDL
NRNFAKHCCEGASSCSSEYCYGLYPEPEPVKAVSFLENNINQIKAYISMHSYS
QHIVFPYSYTRSKSDHELSLVASEAVRALEKISKNTRYTHGHGSETLYLAPGGDD
WIYDLGKYSFTSNPPVKKLLPLSLK"
polyA_site      1546
BASE COUNT      453 a      323 c      297 g      473 t
ORIGIN
Alignment Scores:
Pred. No.:      2,72e-129      Length:      1546
Score:      1327.00      Matches:      244
Percent Similarity:      99.19%      Conservative:      0
Best Local Similarity:      99.19%      Mismatches:      2
Query Match:      99.18%      Indels:      0
DB:      Gaps:      0
US-09-980-881A-4 (1-246) x AB011969 (1-1546)
QY      1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db      362 GCCTCCGCATCGTACTATGACAGTATCACTCACTAAATGAATCTTCTTGGATGAA 421
QY      21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db      422 TTTTAACTGAGAGGCATCTGATATGCTTTACAAAAATCCACATTTGGATCTCTATTGAG 481
QY      41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60

```

```

482 AAGTACCCACTCTATGTTTAAAGGTTTCTGGAAGAAGCAACAGCCCAAAATGCCATA 541
61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
542 TGGATTGACTTGGGAATTCATGCCAGAGAATGGATCTCTCTGCTTTTCTGCTTGTGTTTC 601
81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
602 ATAGGCCATAATCGAATGTGGAAAGAACCGTCTTTCTATCGAACCAATCATTCATC 661
101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
662 GGAACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 721
121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
722 TCCTCATGCTCGAAACCTACTGTGACATTTATTCCTGAGTCAGAACCAAGAGTGAAGGCA 781
141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
782 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAAGATTAAAGCATACATCAGCATGCATTCA 841
161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
842 TACTCCAGCATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGACCATGAG 901
181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
902 GAACTGCTCTAGTAGCAGTGAAGCAGTTCGTGCTATTGAGAAAAATTAGTAAAAATACC 961
201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAsp 220
962 AGGTATACATATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGAGCAT 1021
221 TrpIleTyrAspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeu 240
1022 TGGATCTATGATTTGGGCATCAATATTCGTTTACATCAAAACCCACTCTGAGAGAAGCTT 1081
241 LeuProLeuSerLeuLys 246
1082 TTGCCGCTGCTCTAAAA 1099
RESULT 2
AR232183
LOCUS      AR232183
DEFINITION      Sequence 1 from patent US 6455294.
ACCESSION      AR232183
VERSION      AR232183.1 GI:27274083
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 1625)
AUTHORS      Gan W., Ketchum,K.A., Di Francesco,V. and Beasley,E.M.
TITLE      Isolated human protease proteins, and uses thereof
JOURNAL      Patent: US 6455294-A 1 24-SEP-2002;
FEATURES
Location/Qualifiers
1..1625
/organism="unknown"
BASE COUNT      488 a      334 c      315 g      488 t
ORIGIN
Alignment Scores:
Pred. No.:      2,48e-125      Length:      1625
Score:      1289.50      Matches:      244
Percent Similarity:      92.42%      Conservative:      0
Best Local Similarity:      92.42%      Mismatches:      2
Query Match:      96.38%      Indels:      18
DB:      Gaps:      1
US-09-980-881A-4 (1-246) x AR232183 (1-1625)

```

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
 DB 359 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTCTTGATAGAA 418
 QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
 DB 419 TTTATACTGAGAGCATCTGATATGCTTACAAAATCCACATTTGGATCTCTATTGAG 478
 QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
 DB 479 AAGTACCACCTCTATGTTTAAAGGTTCTGGAAGAAACAAGCAGCAGCAAAATGCCATA 538
 QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
 DB 539 TGGATTGACTGTGGAAATCCATGCGAGAGATGGATCTCTCTGCTTTCTGCTTGTGGTTC 598
 QY 81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
 DB 599 ATAGGCCATATCGAATGTGGAGAAAGACCGTTCTTCTATGGCAACAATCATTTGCATC 658
 QY 101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
 DB 659 GGAACAGACCTGAATAGAACTTTGCTTCCAAACACCTGGTGTGAGGAAGTGCATCCAGT 718
 QY 121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
 DB 719 TCCTCATGCTCGGAACCTACTGTGGACTTTATCTCCTGAGTCAGAACCCAGAAAGTGAAGCA 778
 QY 141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
 DB 779 GTGGTAGTTTCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGCATTC 838
 QY 161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
 DB 839 TACTCCAGCATATAGTGTTCCTATATTCATACAGAAAGTAAAGCAAGACCATGAG 898
 QY 181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
 DB 899 GAACGTGTCTCTAGTAGCAGTGAAGCAGTTCGTGCTATTGAGAAAAATTAGTAAAAATACC 958
 QY 201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspAsp 220
 DB 959 AGGTATACACATGCCATGCTCAGAAACCTTATACCTAGCTCTCTGGAGTGGGAGCAT 1018
 QY 221 TrpIleTyrAspLeuGlyIleLysTyrSerPhe----- 231
 DB 1019 TGGATCTATGTTGGGCATCAATATTCGTTTACAATTGAACTTCGAGATACGGGCACA 1078
 QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLe 243
 DB 1079 TACGGATTCTTGTCCGGAGCGTTACATCAAAACCCACCTGTAGAGAAAGCTTTTGGCGCT 1138
 QY 243 uSerLeuLys 246
 DB 1139 GTCTCTAAAA 1148
 RESULT 3
 LOCUS AR086324
 DEFINITION Sequence 1 from patent US 5985562.
 ACCESSION AR086324
 VERSION AR086324.1 GI:10013090
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.
 1 (bases 1 to 1272)
 Morser, M. John. and Nagashima, M.
 TITLE Method of detecting thrombotic disease risk associated with plasma
 carboxypeptidase B polymorphisms
 JOURNAL Patent: US 5985562-A 1 16-NOV-1999;
 FEATURES Location/Qualifiers

source 1. .1272
 BASE COUNT 375 a 269 c 271 g 357 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,24e-123 Length: 1272
 Score: 1272.00 Matches: 246
 Percent Similarity: 81.73% Conservative: 0
 Best Local Similarity: 81.73% Mismatches: 0
 Query Match: 95.07% Indels: 55
 DB: 6 Gaps: 2
 US-09-980-881A-4 (1-246) x AR086324 (1-1272)
 QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
 DB 343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAATCTATTCTTGATAGAA 402
 QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
 DB 403 TTTATACTGAGAGCATCTGATATGCTTACAAAATCCACATTTGGATCTCTATTGAG 462
 QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
 DB 463 AAGTACCACCTCTATGTTTAAAGGTTCTGGAAGAAACAAGCAGCAGCAAAATGCCATA 522
 QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
 DB 523 TGGATTGACTGTGGATCCATGCCAGAAATGGATCTCTCTGCTTCTGCTTGTGGTTC 582
 QY 81 IleGlyHis----- 83
 DB 583 ATAGGCCATATACTCAATTTCTATGGCAATAGGGCAATATACCAATCTCTCTGAGGCTT 642
 QY 83 ----- 83
 DB 643 GTGGATTCTATGTTATGCCGGTGGTTAATGTGCGCGTTATGACTACTCATGGAAGAAAG 702
 QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
 DB 703 AATCGATGTGGAGAAAGAACCGTCTTCTATGCGAACATCATTTGCATCGGACAGAC 762
 QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
 DB 763 CTGAATAGGAACCTTGTCTTCCAAACACTGTGTGAGAGAGTGCATCCAGTCTCTCATGC 822
 QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
 DB 823 TCGGAACCTTACTGTGGACTTTATCTGAGTCAAGAACCAAGAGTGAAGGAGTGGCTAGT 882
 QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
 DB 883 TTCTTGAGAGAAATATCAACAGATTAAGCATACATCAGCATGATCATCTACTCCAG 942
 QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluLeuSer 183
 DB 943 CATATAGTGTTCATATTCCTATACAGAAAGTAAAGCAAGAACCATGAGGAACCTGTCT 1002
 QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
 DB 1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACAGGATATACA 1062
 QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyLysAspTrpIleTyr 223
 DB 1063 CATGGCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGTGGGAGCATTTGATCTAT 1122
 QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
 DB 1123 GATTTGGGCATCAAAATATTCTGTTTACAATTGAATTCGAGATACGGGCACATACGGATTC 1182
 QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246

Db 1183 TTGCTGCCGAGGCTTACATCAAAACCACCTGTAGAGAAGCTTTTGGCGGCTGTCTCTTAAA 1242
QY 246 s 246
Db 1243 A 1243

RESULT 4
LOCUS BD106653 1272 bp DNA linear PAT 18-SEP-2002
DEFINITION Method of detecting thrombotic disease risk.
ACCESSION BD106653
VERSION BD106653.1 GI:23201471
KEYWORDS JP 2002502253-A/1.
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Morser, M.J. and Nagashima, M.
TITLE Method of detecting thrombotic disease risk
JOURNAL Patent: JP 2002502253-A 1 22-JAN-2002;
SCHERING AG
COMMENT PN JP 2002502253-A/1
PD 22-JAN-2002
PF 02-JUN-1998 JP 1999501465
PR 03-JUN-1997 US 08/869057
PI MICHAEL JOHN MORSE, MARIKO NAGASHIMA
PC C12Q1/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
source 1..1272 Location/Qualifiers
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN

Alignment Scores:
Pred. No.: 1.24e-123 Length: 1272
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-4 (1-246) x BD106653 (1-1272)

QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 343 GCTCCGCGATCGTACTATGACAGTATCATCTCAATGAATCTATCTTGGTAGAA 402

QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 403 TTTATAACTGAGAGGCATCTGATATGCTTACAAAATCCACATTCGATCTCATTTGAG 462

QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 463 ARGATACCCTCTATGTTTAAAGGTTCTGGAAAAGAACAAACAGCAAAATGCCATA 522

QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 523 TGGATTGACTGTGAATCCATGCCAGAGATGATCTCTCTGCTTCTGCTTGTGGTTC 582

QY 81 IleGlyHis----- 83
Db 583 ATAGGCCATATACTCAATCTATGGGATAATAGGCAATATACCAATCTCCTGAGGCTT 642

QY 83 ----- 83
Db 643 GTGGATTCTTATGTTATGCGCGGTGGTGAATGTGGACGGTTATGACTACTCATCATGAGAAAAG 702

QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103

Db 703 AATCGAATGTGAGAAAGAACCGTTCTTTCTATGGAACAATCATTTGTCATCGGAACAGAC 762
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 763 CTGAATAGGAACCTTTGCTTCCAAACACTGGTGTGAGAAAGGTCATCCAGTTCTCATGC 822

QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 823 TCGGAACCTACTGTGACTTTTATCTGAGTCAGAACCAAGAGTGAAGCAGTGGCTAGT 882

QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 883 TTCTTCGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCTCATCTCCAG 942

QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 193
Db 943 CATATAGTGTTCCTATATTCCTATATACGAAGTAAAGCAAGAACCATGAGGAACCTGTCT 1002

QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1003 CTAGTAGCCAGTGAACAGTTCTGCTATTGAGAAAACCTAGTAAATACAGGTATACA 1062

QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
Db 1063 CATGGCCATGGCTCAGAAAACCTTATACCTAGCTCCTGGAGGTGGGACCATGATCTAT 1122

QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 1123 GATTTGGGCATCAAAATATTCGTTTACAAATTGCAATTCGAGATACGGGCACATACGGATTC 1182

QY 232 -----ThrSerAsnProValGluLysLeuProLeuSerLeuLys 246
Db 1183 TTGCTCCGAGGCTTACATCAAAACCCACCTGTAGAGAAGCTTTTCCCGCTGTCTTAAA 1242

QY 246 s 246
Db 1243 A 1243

RESULT 5
LOCUS AX409750 1728 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 2397 from Patent WO0229103.
ACCESSION AX409750
VERSION AX409750.1 GI:21442455
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 2397 11-APR-2002;
GENE LOGIC INC (US)
FEATURES Location/Qualifiers
source 1..1728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="ENBL/GenBank Accession No. M75106"

BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN

Alignment Scores:
Pred. No.: 1.84e-123 Length: 1728
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-4 (1-246) x AX409750 (1-1728)

```
Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 362 GCCTCCGCATCGTACTGATGACAGTATCACTCAATAATGAAATCTATTCTTGGATAGAA 421
Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 422 TTATTAACCTGAGAGGATCGTGTATGCTTCAAAAATCCAAATCGATTCCTCATTTGAG 481
Qy 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAlaIle 60
Db 482 AAGTACCCACTCTATGTTTAAAGGTTTCTGGAAAAAGAACAAACAGCCAAAAATGCCATA 541
Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTrpPhe 80
Db 542 TGGATTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGGTTC 601
Qy 81 IleGlyHis----- 83
Db 602 ATAGGCCATATAACTCAATTCTATGGGTAATAGGGCAATATACCAATCTCTGAGGCTT 661
Qy 83 ----- 83
Db 662 GTGGATTTCATGTTATGCGGTGTTAATGTGACGGTTATGACTACTCATCGAAAAAG 721
Qy 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 722 AATCGAATGTGGAGAAAGACCGTCTTCTATCGCAACAATCATTTGCATCGAACAGAC 781
Qy 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluClyAlaSerSerSerCys 123
Db 782 CTGAATAGGAACCTTGTCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTCTCTATGC 841
Qy 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 842 TCGGAACCTACTGTGGACTTATCTGAGTCAAGAACAGAAAGTGAAGCAGTGGCTAGT 901
Qy 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 902 TTCTTGAGAAAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCATCTCCAG 961
Qy 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 962 CATATAGTGTTCATATTCCTATACAGCAAGTAAAGCAACCAACCATGAGGAAGTGTCT 1021
Qy 184 LeuValAlaSerGluAlaValAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081
Qy 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyr 223
Db 1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGACGATTCGATCTAT 1141
Qy 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 1142 GATTGGGCATCAAAATATTCGTTTACAATTGAACCTTCGAGATACGGGCACATACGGATT 1201
Qy 232 -----ThrsSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
Db 1202 TTGCTCCGGAGCGTTACATCAAAACCCACCTGTAGAGAGCTTTTGCCCGCTGTCTCTAAA 1261
Qy 246 s 246
Db 1262 A 1262
```

RESULT 6
AX706330
LOCUS
DEFINITION
Sequence 1 from Patent WO03014391.
AX706330
ACCESSION
VERSION
AX706330.1 GI:29562754
KEYWORDS
SOURCE
Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Morten, J.E. and Jackson, K.N.

AUTHORS

Method Patent: WO 03014391-A 1 20-FEB-2003;

TITLE

Astrazeneca AB (SE)

JOURNAL

Location/Qualifiers

1. .1728

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

FEATURES

BASE COUNT 518 a 354 c 338 g 518 t

ORIGIN

Alignment Scores:

Pred. No.: 1,84e-123 Length: 1728

Score: 1272.00 Matches: 246

Percent Similarity: 81.73% Conservative: 0

Best Local Similarity: 81.73% Mismatches: 0

Query Match: 95.07% Indels: 55

DB: 6 Gaps: 2

US-09-980-881A-4 (1-246) x AX706330 (1-1728)

Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20

Db 362 GCCTCCGCATCGTACTGATGACAGTATCACTCAATAATGAAATCTATTCTTGGATAGAA 421

Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

Db 422 TTATTAACCTGAGAGGATCGTGTATGCTTCAAAAATCCAAATCGATTCCTCATTTGAG 481

Qy 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAlaIle 60

Db 482 AAGTACCCACTCTATGTTTAAAGGTTTCTGGAAAAAGAACAAACAGCCAAAAATGCCATA 541

Qy 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTrpPhe 80

Db 542 TGGATTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTGCTTTCTGCTTGTGGTTC 601

Qy 81 IleGlyHis----- 83

Db 602 ATAGGCCATATAACTCAATTCTATGGGTAATAGGGCAATATACCAATCTCTGAGGCTT 661

Qy 83 ----- 83

Db 662 GTGGATTTCATGTTATGCGGTGTTAATGTGACGGTTATGACTACTCATCGAAAAAG 721

Qy 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103

Db 722 AATCGAATGTGGAGAAAGACCGTCTTCTATCGCAACAATCATTTGCATCGAACAGAC 781

Qy 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123

Db 782 CTGAATAGGAACCTTGTCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTCTCTCATGC 841

Qy 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143

Db 842 TCGGAACCTACTGTGGACTTATCTGAGTCAAGAACCAACCAAGTGAAGCAGTGGGTAGT 901

Qy 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163

Db 902 TTCTTGAGAAAGAAATATCAACAGATTAAGCATACATCAGCATGCATTCATCTACTCCAG 961

Qy 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183

Db 962 CATATAGTGTTCATATTCCTATACAGCAAGTAAAGCAACCAACCATGAGGAAGTGTCT 1021

Qy 184 LeuValAlaSerGluAlaValAlaIleGluLysThrSerLysAsnThrArgTyrThr 203

Db 1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081

```
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrrPilleTyr 223
Db 1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGACGATGGAICTAT 1141
QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 1142 GATTTGGGCATCAAAATATTCGTTTACAAATTCGAGATTCAGGGCACATACGGATTC 1201
QY 232 -----ThrSerAsnProProValGluLeuLeuProLeuSerLeuLeu 246
Db 1202 TTGCTGCCGAGCGCTTACATCAAAACCCACCTGTAGAGAAGCTTTTGGCCGTGTCTCTAAA 1261
QY 246 s 246
Db 1262 A 1262

RESULT 7
HUMPCPBX 1728 bp mRNA linear PRI 07-JAN-1995
LOCUS Human prepro-plasma carboxypeptidase B mRNA, complete cds.
DEFINITION M75106
ACCESSION M75106.1 GI:189686
KEYWORDS plasma carboxypeptidase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE Isolation, molecular cloning, and partial characterization of a
novel carboxypeptidase B from human plasma
JOURNAL J. Biol. Chem. 266 (32), 21833-21838 (1991)
MEDLINE 92042093
PUBMED 1939207
COMMENT source text: Homo sapiens liver cDNA to mRNA.
FEATURES
Location/Qualifiers
1..1728
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="liver"
1..1728
/genes="pCPB"
20..1291
/genes="pCPB"
/notes="prepro-plasma carboxypeptidase B"
/codon_start=1
/protein_id="AAA60042.1"
/db_xref="GI:189687"
/translation="MKLCSLAVLVPIVLFCQHVFAFQSGQVLALPRTSRQVQVLQN
LTTTIEIVLWQPVTDLI VKKQVHFVNASDVNDVKAHLNVGSPICSVLLADVDLI
QQQISNDTSPRASASYEQYHSNIEYFIEFIERHPDMLTKIHIGSSFEKYPFLY
LKVSKEQTAQNAIWDICGTHAREWISPAFLWFIHQFYIGIQTNNLRVDFY
VMPVNVGDYDSKNNRWRKNSFYANNHCIGTDLNRNFASKHWCBSGSSCSSE
TYCLYPSSEPEYKAVASFLRRNIQKAVISMHSYSOHLVPEYSYTRSKSDHEELS
LVASEAVRALEKTSKNTRYTHGSETLYLAPCGGDDMIYDLGIKISPTIELRDTGY
GFULPRYIKTPCREAPAAVSKIAWHVRNV"
mat_peptide
86..1288
/genes="pCPB"
/product="plasma carboxypeptidase B"
BASE COUNT 518 a 354 c 338 g 518 t
ORIGIN

Alignment Scores:
Pred. No.: 1.84e-123 Length: 1728
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 9 Gaps: 2

US-09-980-881A-4 (1-246) x HUMPCPBX (1-1728)
```

```
QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrrPilleGlu 20
Db 362 GCCTCCGCATCGTACTATGAAACAGTATCACTCATAAATGAAATCTATTCTTGGATAGAA 421
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 422 TTTATAACTGAGAGGCATCTCGATATGCTTACAAAAATCCACATTCGATCTCATTTGAG 481
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyGluGlnThrAlaLysAsnAlaIle 60
Db 482 AAGTACCCCATCTATGTTTAAAGGTTCTGGAAAAAGAAACAAACAGCCAAAAATGCCATA 541
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrrPilleSerProAlaPheCysLeuTrrPhe 80
Db 542 TGGATTGACTGTGGAATCCATGCCAGAAATGATCTCTCTGCTTTCTGCTTGTGGTTC 601
QY 81 IleGlyHis----- 83
Db 602 ATAGGCCATATAACTCAATTCATGGGATAATAGGGCAATATACCAATCTCTCTGAGGCTT 661
QY 83 ----- 83
Db 662 GTGGATTTCTATGTTATGCCGGTGGTAAATGTGACGGTTATGACTACTCATGGAAAAAG 721
QY 84 AsnArgMetTrrArgLysAsnArgSerPheTrrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 722 AATCGAATGTGGAAGAAAGAACCGCTCTTCTATCGCAACAATCATTCGATCGGAACAGAC 781
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrrCysGluGluGlyAlaSerSerSerCys 123
Db 782 CTGAATAGGAACCTTTGCTTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCCTCATGC 841
QY 124 SerGluThrTrrCysGlyLeuTrrProGluSerGluProGluValLysAlaValAlaSer 143
Db 842 TCGGAACCTTACTGTGGAAGTCTTATCTGAGTCAGAACCAAGTGAAGCGAGTGGCTAGT 901
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTrrIleSerMetHisSerTrrSerGln 163
Db 902 TTTCTGGAAGAAATATCAACAGATTAAGCATATACATCAGCATGCAATTCATCTCCAG 961
QY 164 HisIleValPheProTrrSerTrrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 962 CATATAGTGTTCATATTCCTATATACCAAGATAAAGCAAGACCATCAGGAACGTGCT 1021
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTrrThr 203
Db 1022 CTAGTAGCCAGTCAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081
QY 204 HisGlyHisGlySerGluThrLeuTrrLeuAlaProGlyGlyGlyAspTrrPilleTyr 223
Db 1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGACGATGGAICTAT 1141
QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 1142 GATTTGGGCATCAAAATATTCGTTTACAAATTCGAGATTCAGGGCACATACGGATTC 1201
QY 232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLeu 246
Db 1202 TTGCTGCCGAGCGTTCATCAAAACCCACCTGTAGAGAAGCTTTTGGCCGTGTCTCTAAA 1261
QY 246 s 246
Db 1262 A 1262

RESULT 8
LOCUS I16100
DEFINITION Sequence 2 from patent US 5474901.
ACCESSION I16100
VERSION I16100.1 GI:1251008
KEYWORDS Unknown.
```

```
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Antibodies to human carboxypeptidase B and methods of use thereof
JOURNAL Patent: US 5474901-A 2 12-DEC-1995;
FEATURES
    source
        1..1749
            /organism="unknown"
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN

Alignment Scores:
Pred. No.: 1,87e-123 Length: 1749
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-4 (1-246) x I16100 (1-1749)
QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLysTyrSerTrpIleGlu 20
DB 383 GCCTCGCATCGTACTATGACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAA 442
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 443 TTTATAACTGAGAGGCATCTGATATGCTTACAAAATCCACATTCGATCCTCATTTGAG 502
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 503 AAGTACCCACTCTATGTTTAAAGGTTCTGGAAGAAGAACCAACCCAAAATGCGCATA 562
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTrpPhe 80
DB 563 TGGATTGACTGTGGAATCCATGCCAGAGTGGATCTCTCTGCTTCTGCTTGTGTTGCTTC 622
QY 81 IleGlyHis----- 83
DB 623 ATAGGCCATATAACTCAATTTCTATGGGCAATATAGGCAATATACCAATCTCTCTGAGGCTT 682
QY 83 ----- 83
DB 683 GTGGATTTCTATGTTATGCGGTGTTAATGTGGACGGTTATGACTACTACTATCGGAAAAAG 742
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
DB 743 AATCGAATGTGGAAGAAGAACCGCTTCTTATGCGAACAATCATTTGCATCGGAACAGAC 802
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
DB 803 CTGAATAGGACTTTGCTTCCAAACACTGGTGTGAGGAAGTGATCCAGTTCCTCATGCG 862
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
DB 863 TCGGAACCTACTGTGGAATTTATCTCTGAGTCCAGAACCCAGAGTGAAGCGAGTGGTAGT 922
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMethIleSerTyrSerGln 163
DB 923 TTCTTGAGAAGAAATATCAACCAAGATTAAGCATATACATGAGCATGTCATCTATCTCCAG 982
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
DB 983 CATATAGTGTTCATATTCCTATACACGAGTAAAGCAAGCAACCATGAGGACTGTCT 1042
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
DB 1043 CTAGTAGGCAGTGAAGCAGTTCTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyCysAspTrpIleTyr 223
DB 1103 CATGGCCATGGCTCAGAAAACCTTATACCTAGCTCTCTGGAGGTGGGACGATTTGGATCTAT 1162

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1749)
AUTHORS Drayna,D.T. and Eaton,D.L.
TITLE Antibodies to human carboxypeptidase B and methods of use thereof
JOURNAL Patent: US 5474901-A 2 12-DEC-1995;
FEATURES
    source
        1..1749
            /organism="unknown"
BASE COUNT 521 a 361 c 342 g 525 t
ORIGIN

Alignment Scores:
Pred. No.: 1,87e-123 Length: 1749
Score: 1272.00 Matches: 246
Percent Similarity: 81.73% Conservative: 0
Best Local Similarity: 81.73% Mismatches: 0
Query Match: 95.07% Indels: 55
DB: 6 Gaps: 2

US-09-980-881A-4 (1-246) x I16100 (1-1749)
QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluLysTyrSerTrpIleGlu 20
DB 383 GCCTCGCATCGTACTATGACAGTATCACTCACTAAATGAATCTATTCTTGGATAGAA 442
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
DB 443 TTTATAACTGAGAGGCATCTGATATGCTTACAAAATCCACATTCGATCCTCATTTGAG 502
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
DB 503 AAGTACCCACTCTATGTTTAAAGGTTCTGGAAGAAGAACCAACCCAAAATGCGCATA 562
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTrpPhe 80
DB 563 TGGATTGACTGTGGAATCCATGCCAGAGTGGATCTCTCTGCTTCTGCTTGTGTTGCTTC 622
QY 81 IleGlyHis----- 83
DB 623 ATAGGCCATATAACTCAATTTCTATGGGCAATATAGGCAATATACCAATCTCTCTGAGGCTT 682
QY 83 ----- 83
DB 683 GTGGATTTCTATGTTATGCGGTGTTAATGTGGACGGTTATGACTACTACTATCGGAAAAAG 742
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
DB 743 AATCGAATGTGGAAGAAGAACCGCTTCTTATGCGAACAATCATTTGCATCGGAACAGAC 802
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
DB 803 CTGAATAGGACTTTGCTTCCAAACACTGGTGTGAGGAAGTGATCCAGTTCCTCATGCG 862
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
DB 863 TCGGAACCTACTGTGGAATTTATCTCTGAGTCCAGAACCCAGAGTGAAGCGAGTGGTAGT 922
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMethIleSerTyrSerGln 163
DB 923 TTCTTGAGAAGAAATATCAACCAAGATTAAGCATATACATGAGCATGTCATCTATCTCCAG 982
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
DB 983 CATATAGTGTTCATATTCCTATACACGAGTAAAGCAAGCAACCATGAGGACTGTCT 1042
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
DB 1043 CTAGTAGGCAGTGAAGCAGTTCTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyCysAspTrpIleTyr 223
DB 1103 CATGGCCATGGCTCAGAAAACCTTATACCTAGCTCTCTGGAGGTGGGACGATTTGGATCTAT 1162
```



```

Db      803 CTGAATAGGAACCTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTCTCCTCATGC 862
Qy      124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db      863 TCGGAACCTTACTGTGAGCTTTATCTGAGTCAGAACCAAGAGTGAAGGAGTGGCTAGT 922
Qy      144 PheLeuArgArgAsnIleAenGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db      923 TTCTTGAGAGAATATACACACAGATTAAAGCATACATCAGCATGCATTCATACATCCAG 982
Qy      164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db      983 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGCAACCATCAGGAACGTCT 1042
Qy      184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db      1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAACCTAGTAAATACCAAGTATACA 1102
Qy      204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTyrPheTyr 223
Db      1103 CATGGCCATGGCTCAGAACCTTATACCTAGCTCTGGAGGTGGGACGATTGGATCTAT 1162
Qy      224 AspLeuGlyIleLysTyrSerPhe----- 231
Db      1163 GATTGGGCATCAATATTCGTTTACAAATTCGAATTCGAGATACGGGCACATACGGATTC 1222
Qy      232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
Db      1223 TTCTGCGCGAGCGTTACATCAACCCACCTGTAGAGAACCTTTTCCCGCTGTCTTAA 1282
Qy      246 s 246
Db      1283 A 1283

RESULT 10
LOCUS   BT006936
DEFINITION Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA, complete cds.
ACCESSION BT006936
VERSION   BT006936.1 GI:30582710
KEYWORDS  FLI CDNA.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 1272)
AUTHORS   Kalline,N., Chen,X., Rolfes,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE     Cloning of human full-length cDNAs in BD Creator(TM) System Donor vector
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 1272)
AUTHORS   Kalline,N., Chen,X., Rolfes,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE     Direct Submission
JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT   This cDNA clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each cDNA has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The cDNA has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame.
FEATURES
source    Location/Qualifiers
          1..1272
            /organism="Homo sapiens"

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CH00156X1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC collection"
/lab_host="DH5alpha T1 resistant"
/notes="Vector: pDNR-Dual"
1..1272
/codon_start=1
/product="carboxypeptidase B2 (plasma, carboxypeptidase U)"
/protein_id="AAP35582.1"
/db_xref="GI:30582711"
/translation="MKLCSLAVLPVLFCQHVFAFGQGVLAALPRTSRQVQVLQN
LTTTEIVLWQPTADLIYKKQVHPFNASVDNVKHLNLSGIPCSVLLADVEDLI
QQQINDTVSPRASAIYEQYHSLNEIYSEIFITERHDMUTKIHIGSFYPIYIV
LKVSQKQAKNAIWDICGIHAREWISPAFCWIFGHITQFQIGIQYTNLRVLDFY
VNVVNDGYDYSWKNRMRNRSFPANNHCGITDLENFASGKHWCEGASSSSE
TYCVGLYPEPEPVKAVASFLRNINQIKAYISMHSYSHIVFPYSYTRSKSDHELS
LVASAVRAIEKTSKNTRYTHGSGSETLVLPGGDDDIYDLGIKYSFIELRDTCTY
GFLLPERYIKPTCREAFAPAAVSKIAHVIRNV"
BASE COUNT 373 a 268 c 273 g 358 t
ORIGIN

Alignment Scores:
Pred. No.: 4,14e-123 Length: 1272
Score: 1267.00 Matches: 245
Percent Similarity: 81.40% Conservative: 0
Best Local Similarity: 81.40% Mismatches: 1
Query Match: 94.69% Indels: 55
DB: 9 Gaps: 2

US-09-980-881A-4 (1-246) x BT006936 (1-1272)

Qy      1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPheGlu 20
Db      343 GCCTCGCATCGTACTATGACAGTATCACTCACTAAATGAAATCTATTCTTGATAGAA 402
Qy      21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db      403 TTTATAACTGAGAGGCATCTGATATGCTTACAAAATCCACATTGGATCTCTCATTTGAG 462
Qy      41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db      463 AAGTACCCACTCTATGTTTAAAGGTTTCUGAAAAGAACACAGCACCAAAAATGCCATA 522
Qy      61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTyrPhe 80
Db      523 TGGATTGACTGTGAATCCATGCCAGAGATGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 582
Qy      81 IleGlyHis----- 83
Db      583 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCTCTGAGCTT 642
Qy      83 ----- 83
Db      643 GTGGATTCTATGTTATGCGGGTGGTTAATGTGGATGGTTATGACTACTCATCGAAAAAAG 702
Qy      84 AsnArgMetTyrArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db      703 AATCGAATGGGAGAAAGAACCGTCTTCTTCTATCGCAACAATCATTTGCATCGGAACAGAC 762
Qy      104 LeuAsnArgAsnPheAlaSerLysHisTyrCysGluGluGluLysAlaSerSerSerCys 123
Db      763 CTGAATAGGAACCTTGTCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTCTCCTCATGC 822
Qy      124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db      823 TCGGAACCTTACTGTGAGCTTTATCTGAGTCAGAACCAAGAGTGAAGGAGTGGCTAGT 882
Qy      144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db      883 TTCTTGAGAGAATATACCAACAGATTAAAGCATACATCAGCATGCATTCATCATCTCCAG 942

```

```

QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 943 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGCAACCATGAGGAAGTCTCT 1002
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1003 CTAGTAGCCAGTACGAGTTCGTCTGCTATTGAGAAACTAGTAAATAACACAGTATACA 1062
QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTyrPheTyr 223
Db 1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGTGGGACGATGGATCTAT 1122
QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
Db 1123 GATTGGGCATCAAAATATTCGTTTACAAATTGAACCTTCGAGATACGGGCACATACGGATTC 1182
QY 232 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLys 246
Db 1183 TTGCTCGCGGAGGCTTACATCAAAACCCACCTGTAGAGAGCTTTTGGCGTGTCTTAAA 1242
QY 246 S 246
Db 1243 A 1243

RESULT 11
BT007990
LOCUS 1272 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
carboxypeptidase U) mRNA, partial cds.
ACCESSION BT007990
VERSION BT007990.1 GI:30584818
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM artificial construct
REFERENCE 1 (bases 1 to 1272)
AUTHORS Kalline,N., Chen,X., Rolfes,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1272)
AUTHORS Kalline,N., Chen,X., Rolfes,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the Sali and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after Sali site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
Location/Qualifiers
1. 1272
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH001561.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/notes="Vector: pDNR-Dual"
1. >1272
/notes="Mutations: 1271.Stop->Leu"
/codon_start=1

CDS

```

```

/transl_table=11
/product="Homo sapiens carboxypeptidase B2 (plasma,
carboxypeptidase U)"
/protein_id="AAP36662.1"
/db_xref="GI:30584819"
/translation="MKLCSLAVLPVTLFCQHVPAFGQGVLAALPRTSRQVQLQN
LTTTVEILWQPVTDLIYKKQVHFFVNASDVNVKAHLNVSGIPCSVLADVEDLI
QQQISNDTSPRASASYEQYHSNLNIEYSWIEFITERHFDMLTKIHGSSFEKPLVY
LVKSGKEQAANAIDDCGIHAREWISPAFLWFIQHTQFYGIIGQYNLLRLVDY
LVNVVDGDYDYSWKKNRMRKNRSFYANNHCIGTDLNRNFASKHCEGASSCSSE
TYCGLYPESEPKAVASFLRNINQIKAYISMHSYQHVIFPFPYSYTRSKSKDHEELS
LVASEAVRAIEKTSKNTYRTHGSHSTLVLPAGGGDDWIYDLGIKYSFTELRDTGY
GELLPERYIKPTCREAFAAVSKIAHWIRNVL"
BASE COUNT 372 a 268 c 273 g 359 t
ORIGIN
Alignment Scores:
Pred. No.: 4,14e-123 Length: 1272
Score: 1267.00 Matches: 245
Percent Similarity: 81.40% Conservative: 0
Best Local Similarity: 81.40% Mismatches: 1
Query Match: 94.69% Indels: 55
DB: 12 Gaps: 2
US-09-980-881A-4 (1-246) x BT007990 (1-1272)
QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTyrPheGlu 20
Db 343 GCCTCGGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGATAGAA 402
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
Db 403 TTTATAAAGTGGAGGATCTGATATGCTTACAAAAATCCACATTCGATCTCTATTGAG 462
QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 463 AGTACCCACTCTATGTTTAAAGGTTTCTGAAAAAGAACACGACGCAAAATATGCCATA 522
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrPheSerProAlaPheCysLeuTyrPhe 80
Db 523 TGGATTGACTGTGGAATCCATGCCAGAGATGGATCTCTCTCTCTCTCTCTCTCTCTCTCT 582
QY 81 IleGlyHis----- 83
Db 583 ATAGGCCATATAACTCAATTTCTATGGGCAATAATAGGCAATATACCAATCTCTCTGAGGCTT 642
QY 83 ----- 83
Db 643 GTGGATTTCTATGTTATGCGGGTGGTGTATGTGGATGGTTATGACTACTCATCGAAAAAG 702
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
Db 703 AATCGAATGGGAAAGAACCGTCTTCTTATCGCAACAATCATTCGATCGGACAGAC 762
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 763 CTGAATAGGAACCTTCTCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTCTCTCATGC 822
QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
Db 823 TCGGAACCTTACTGTGGACTTTTATCTGAGTTCAGAACCAAGAGGAGGAGGAGTGTAGT 882
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
Db 883 TTCTTGAGGAAGAAATATCAACAGATTAAGCATATACATCAGCATGATCATCTACTCCAG 942
QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
Db 943 CATATAGTGTTCATATTCCTATACACGAAGTAAAGCAAGCAACCATGAGGAAGTCTCT 1002
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
Db 1003 CTAGTAGCCAGTACGAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAAGTATACA 1062

```

QY 204 HisGlyHisGlySerGluThrLeuTyLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
 |||||
 Db 1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGACGATTGATCTAT 1122
 |||||
 QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
 |||||
 Db 1123 GATTGGGCATCAAAATATTGTTTACAAATTGAACTTCGAGATACGGGCACATACGGATTC 1182
 |||||
 QY 232 -----ThrSerIleProValGluLysLeuLeuProLeuSerLeuLeu 246
 |||||
 Db 1183 TTGCTGCGGAGGCTTACATCAAAACCCCTGTAGAGAGCTTTTGGCGGTGTCTCTAAA 1242
 |||||
 QY 246 s 246
 |||||
 Db 1243 A 1243

RESULT 12
 BC007057
 LOCUS
 DEFINITION Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
 IMAGE:3934520, mRNA, complete cds.
 ACCESSION BC007057
 VERSION 1
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1715)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: CLONTECH
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNase Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mdc@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 16 Row: f Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4503004.
 FEATURES
 source
 1. .1715
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="LocusID:1361"
 /db_xref="taxon:9606"
 /clone="MGC:12495 IMAGE:3934520"
 /tissue_type="Skeletal Muscle"
 /clone_lib="NIH_MGC_81"
 /lab_host="DH10B"
 /notes="Vector: pDNR-LIB"
 25. .1296
 /codon_start=1
 /product="carboxypeptidase B2 (plasma)"
 /protein_id="AA07057.1"
 /db_xref="GI:13937897"
 /translations="MKLCSLAVLPIVLFCQHVFAFQSGQVLAALPRTSRQVQLQN
 LTTTVEIVLWQPTADLVKKQVHFVNASDVNDVKHLNVSGIPCSVLLADVEDLI
 QQQISNDTVSRASASYEQYHSLNIYSWIEFITERHPDMLTKIHIGSFSEKYPVLYV

CDS

BASE COUNT 518 a 348 c 341 g 508 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,09e-123 Length: 1715
 Score: 1267.00 Matches: 245
 Percent Similarity: 81.40% Conservative: 0
 Best Local Similarity: 81.40% Mismatches: 1
 Query Match: 94.69% Indels: 55
 DB: 9 Gaps: 2
 US-09-980-881A-4 (1-246) x BC007057 (1-1715)
 QY 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
 |||||
 Db 367 GCTCCGCACTGCTATGAACTATCACTCACTAAATGAATCTATTCTTGGATAGAA 426
 |||||
 QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
 |||||
 Db 427 TTTATAACTGAGAGGCATCTGATATGCTTACAAAATCCACATTCGATCCTCATTTGAG 486
 |||||
 QY 41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
 |||||
 Db 487 AAGTACCACCTCTATGTTTAAAGGTTTCTGAAAAGAAACAAGCAGCAAAATATGCCATA 546
 |||||
 QY 61 TrpIleAspCysGlyIleHisAlaArgGluTyrTrpIleSerProAlaPheCysLeuTrpPhe 80
 |||||
 Db 547 TGGATTGACTGTGGAATCCATGCCAGAAATGGATCTCTCTGCTTTCTGCTTGTGTTTC 606
 |||||
 QY 81 IleGlyHis----- 83
 |||||
 Db 607 ATAGGCCATATACTCAATTCATGGGATAATAGGCAATATACCAATCTCTCTGAGGCTT 666
 |||||
 QY 83 ----- 83
 |||||
 Db 667 GTGGATTCTATGTTATGCGGTGTTAATGTGGATGGTTTATGACTACTCATGGAAG 726
 |||||
 QY 84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
 |||||
 Db 727 AATCGAATGTGGAGAAGAACCGTTCTTCTATGCGAACCAATCATTCATCGCAACAGAC 786
 |||||
 QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
 |||||
 Db 787 CTGAATAGGAACCTTCTTCCAAACACTGGTGTGAGGAAGTGCATCCAGTTCTCATGC 846
 |||||
 QY 124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
 |||||
 Db 847 TCGGAACCTTACTGTGGACTTTTATCTGAGTCAGAACCAAGAGTGAAGGAGTGGTGTAGT 906
 |||||
 QY 144 PheLeuArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
 |||||
 Db 907 TTCTTGAGAAGAAATATCAACAGATTAAGCATATACATCAGCATCATCATCTCCAG 966
 |||||
 QY 164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
 |||||
 Db 967 CATATAGTGTTCATATTCCTATATACAGNAAGTAAAGCAAGCAACCATGAGGACTGTCT 1026
 |||||
 QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
 |||||
 Db 1027 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAAGTATACA 1086
 |||||
 QY 204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspTrpIleTyr 223
 |||||
 Db 1087 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCTCTGGAGTGGGACGATGATCTAT 1146
 |||||
 QY 224 AspLeuGlyIleLysTyrSerPhe----- 231
 |||||
 Db 1147 GATTGGGCATCAAAATATTGTTTACAAATTGAATTCGAGATACGGGCACATACGGATTC 1206
 |||||

```

QY 232 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLeu 246
Db 1207 TTGTCGGGAGCGTTATCATCAAAACCCACCTGTAGAGAAGCTTTTGGCGTGTCTCTAAA 1266
QY 246 s 246
Db 1267 A 1267

RESULT 13
LOCUS AB021968 1421 bp mRNA linear ROD 27-JUL-2000
DEFINITION Mus musculus mRNA for carboxypeptidase R, complete cds.
ACCESSION AB021968
VERSION AB021968.1 GI:9558447
KEYWORDS carboxypeptidase R.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Sato, T., Miwa, T., Akatsu, H., Matsukawa, N., Obata, K., Okada, N.,
Campbell, W. and Okada, H.
Pro-carboxypeptidase R is an acute phase protein in the mouse,
whereas carboxypeptidase N is not
J. Immunol. 165 (2), 1053-1058 (2000)
20341711
10878383
PUBMED
REFERENCE 1 (bases 1 to 1421)
AUTHORS Sato, T.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1998) Tomoo Sato, Fukushima Hospital, Choju
Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi
441-8124, Japan (E-mail: tomoo@tcp-ip.or.jp, Tel: 81-532-46-7511,
Fax: 81-532-46-4899)
FEATURES
source Location/Qualifiers
1. 1421
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
17..1285
/codon_start=1
/product="carboxypeptidase R"
/protein_id="BAB03402.1"
/db_xref="GI:9558448"
/translatons="MKLHGILVAILLYEQHGFAFGOSVLSALPRTSRQVOLLQNL
TTTVEVLWQPTAEPTKEKKVEFFVNASDVSVDKHLNVSRIPNVLNMNVEDLIE
QQTENDTVSPRASAYEYQHSLEIYSLWIEVTEQHPDMLQKIYIGSSFEPKPLVL
KVSQKQRIKAIWIDCGIHAREWISPAFCWIFIGVYTPHGENLYTLRLRHVDYFI
MPVMVDGYDITWKNNRMRKNSAHKNNRCVGTDLNRPNSPKHWCCKGSSSCSET
YCGLYPESEPVKAVADFLRRNDHDKAYISMHSYQQILFPYSYNRSKSKHHEELSL
VASEAVRAIESINKNRYTHGSGESLYLAPGSGDDWIDYLGIKYSFTIELRDTGRYG
FLLPERYIKPTCAEALAAISKIVHWIRNT"
polya_site 1421
/notes="11 a nucleotides"
BASE COUNT 418 a 320 c 305 g 378 t
ORIGIN

Alignment Scores:
Pred. No.: 8.64e-107 Length: 1421
Score: 1112.00 Matches: 217
Percent Similarity: 75.75% Conservative: 11
Best Local Similarity: 72.09% Mismatches: 18
Query Match: 83.11% Indels: 55
DB: 10 Gaps: 2

US-09-980-881A-4 (1-246) x AB021968 (1-1421)

QY 1 AlaSerAlaSerTrpTrpGluGlnTrpHisSerLeuAsnGluTleTrpSerTrpIleGlu 20
Db 356 GCTTCGGCTTCATCTATGAGCAGTATCATCGCTAATGAATCATTTCTCGATAGAA 415
QY 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

```

```

Db 416 GTCACTAACTGAACAGCATCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAG 475
QY 41 LysTrpProLeuTrpValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
Db 476 AAGTACCACCTTTATGTTTAAAGGCTCAGGAAGGACAAAGAATCAANAATGCCATC 535
QY 61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
Db 536 TGGATCGACTGTGGAATCCATGCCAGAGATGGATTTCACCTGCTTTCTGTTTGTGCTTC 595
QY 81 IleGlyHis----- 83
Db 596 ATAGGCTACGTGACACAAATTCATGGGAAAGAAATCTGTATACCAGACTTCTTGAGGCAC 655
QY 83 ----- 83
Db 656 GTGGATTCTACATCATGCCCGGTGATGAACGTGGATGGCTATGACTACACAGTGGAAAAG 715
QY 84 AsnArgMetTrpArgLysAsnArgSerPheTrpAlaAsnAsnHisCysIleGlyThrAsp 103
Db 716 AATCGAATGTGAGGAAGAACCGCTCTGCTCACAGAACAACCGCTCGTGGGCACAGAC 775
QY 104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys 123
Db 776 CTGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGTGGCTCAAGTTCCTCTGTC 835
QY 124 SerGluThrTrpCysGlyLeuTrpProGluSerGluProGluValLysAlaValAlaSer 143
Db 836 TCTGAAACCTACTGTGGACTTTATCTGAGCTGAGCCAGAGGTGAAGGAGTGGCTGAC 895
QY 144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTrpIleSerMetHisSerTrpSerGln 163
Db 896 TTCTTGAGAAGAAATATCGACCACATTAAAGCTTTACATCAGTATGCACTCATCTCCCAA 955
QY 164 HisIleValPheProTrpSerTrpThrArgSerLysSerLysPheHisGluGluLeuSer 183
Db 956 CAAATACTGTTCCTCTTCTTATAACAGAAGCAAGCAAGGACCACCAAGAACTGTCT 1015
QY 184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTrpThr 203
Db 1016 CTAGTGGCCAGCAGACAGTTCGTGCAATTTGAAGTATTATTAATAAACACACAGTACCA 1075
QY 204 HisGlyHisGlySerGluThrLeuTrpLeuAlaProGlyGlyGlyAspTrpIleTrp 223
Db 1076 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATGGATCTAT 1135
QY 224 AspLeuGlyIleLysTrpSerPhe----- 231
Db 1136 GATTTGGGCATCAAAATATTCTGTTTACAATTTAGCTCCGAGATACAGGCAGATACGGATT 1195
QY 232 -----ThrSerAsnProValGluLysLeuLeuProLeuSerLeuLeu 246
Db 1196 TTGCTGCCTGAGAGATACATCAAAACCCACTTTGTGCAGAGCTTTGGCCGCCATCTCTAAA 1255
QY 246 s 246
Db 1256 A 1256

RESULT 14
AF164524
LOCUS AF164524
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Taf1)
mRNA, complete cds.
ACCESSION AF164524
VERSION AF164524.1 GI:7416966
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1490)
REFERENCE 1
AUTHORS Marx, P.F., Wagenaar, G.T., Reijerkerk, A., Tiekstra, M.J., van

```

Db	783	AATCGAATGTGGAGGAAGAACCCGCTCTGCTCAACAAGAACACCCGCTGCGTGGGCACAGAC	842
Qy	104	LeuAsnArgAsnPheAlaSerLysHisIstrpCysGluGluGlyAlaSerSerSerCys	123
Db	843	CTGAAACAGGACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGGCTCAAGTTCTCTCTGC	902
Qy	124	SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer	143
Db	903	TCTGAAACCTACTCTGTGGACTTTATCTGAGTCTGAGCCAGAGGTGAAGGCGGTGC	962
Qy	144	PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln	163
Db	963	TTCTTGAGAAAGAATATCGAACCATTTAAAGCTTTATCATGATGCACTCATATCTCCAA	1022
Qy	164	HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer	183
Db	1023	CAATACTGTTCCCTATTCTCTATACAGNAGCAAGACCAAGACCAAGCAAGAACTGTCT	1082
Qy	184	LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr	203
Db	1083	CTAGTGGCCAGCAAGCAGTTCGTGCAATTTGAAGATTATTAATAAAACACCAAGGTACACA	1142
Qy	204	HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr	223
Db	1143	CACGGCAGTGGCTTCAGAAAGTTTATATCTAGCTCTCGAGGTTCTCAGCAATTGGATCTAT	1202
Qy	224	AspLeuGlyIleLysTyrSerPhe-----	231
Db	1203	GAITGGCCATCAAAATATTGTTTTACANTGACTCGAGATACAGCAGATACGGATTC	1262
Qy	232	-----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys	246
Db	1263	TTGCTGCTGAGAGATACATCAAAACCCACTTGTGTCAGAGCTTTGGCGCCACTCTCTAAA	1322
Qy	246	S 246	
Db	1323	A 1323	
RESULT 15			
AF186188			
LOCUS	AF186188	1430 bp mRNA linear	ROD 01-OCT-1999
DEFINITION	Mus musculus carboxypeptidase U mRNA, complete cds.		
ACCESSION	AF186188		
VERSION	AF186188.1	GI:6003651	
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1430)		
AUTHORS	He, Y.C. and Broze, G.		
TITLE	Isolation and characterization of mouse liver carboxypeptidase B gene		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1430)		
AUTHORS	He, Y.C. and Broze, G.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-SEP-1999) Division of Hematology, Washington University School of Medicine, 216 S. Kingshighway, St. Louis, MO 63110, USA		
FEATURES	Location/Qualifiers		
source	1..1430		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="B6/CBAF1J"		
	/db_xref="taxon:10090"		
	20..1288		
	/note="pro-carboxypeptidase-B; thrombin-activatable fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds plasminogen"		
	/codon_start=1		
	/product="carboxypeptidase U"		
	/protein_id="AAAF00528.1"		
CDS			

/db xref="GI:6003652"		/trãnlãtiã="MKLHGLILVAILLYEQHGFARQSGVLSALPRTSRQVOLLQNL	
TTTYVLLQPVTAETFEIKKEVEHFFVNASDVDSVAHLNVSRPFPVLMNVNVEDDIE		224 AspLeuGlyIleIleYsTyrSerPhe-----	
QOTFNDTVSPRASAYEQYSHLNEIYSWIEVLTEQHPDMLQKIYIGSSPEKYPLVYL		1139 GATTTGGGCATCAAAATATTCTGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTTC	
KVSQKQRIKNAIWDICGHAREWISPAFLCWIFGYVTFPHGKENLYTRLLRHVDFTYI		232 -----ThrSerAsnProValGluIleLeuLeuProLeuSerLeuLeu	
MPWNYDGYDTWKNNRMRKNRCSAHNNRCVGTDLNRNPFASKHWCCKGSSSSCSET		1199 TTGCTCGCTGAGAGATACATCAAACTTTGTGCGAAGAGCTTTGGCCGCCCATCTCTAA	
YCLLYPESPEVKADFLRRNDTHIKAVISMHSYSQQILFPVSYNRSKSDHEELSL		246 S 246	
VASEAVRAIGSFENKNTYTHGSSSESLYLAPGSDDDWIDLGIKISFTIELRDTGRYG		1259 A 1259	
FLLPERYIKPTCEALAAALSKIVVHVIRNT"			
BASE COUNT	421 a 320 c 310 g 379 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	9.74e-106	Length:	1430
Score:	1102.00	Matches:	215
Percent Similarity:	75.42%	Conservative:	12
Best Local Similarity:	71.43%	Mismatches:	19
Query Match:	82.36%	Indels:	55
DB:	10	Gaps:	2
US-09-980-881A-4 (1-246) x AF186188 (1-1430)			
QY	1	AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu	20
DB	359	GCCTCCGCTTCATCTATGAGCAGATATCACTCGCTAAATGAATCTATTCCTGGATAGAA	418
QY	21	PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu	40
DB	419	GTCTTAACGAACAGCATCTGACATGCTCCAGAAAATCTACATCGATCATCTTCGAG	478
QY	41	LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle	60
DB	479	AAGTACCCACTTTATGTTTAAAGGCTCTCAGGAAGGAACAAAGAAATCAAAAATGCCATC	538
QY	61	TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe	80
DB	539	TGATTCGACTGTGGAATCCATGCGCAGAAATGGAATTCACCTGCTTTCTGTGTGTGTTTC	598
QY	81	IleGlyHis-----	83
DB	599	ATAGGCTACGTGACACAAATCCATCCATGGAAAGAAAATCTGTATACCAGACTTCTGAGGCAC	658
QY	83	-----	83
DB	659	GTGGATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAG	718
QY	84	AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp	103
DB	719	AATCGAATGTGGAGGAAGAACCCGCTCTGCTCAAGAACCAACCCGCTGCGTGGGCACAGAC	778
QY	104	LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCys	123
DB	779	CTGAACAGGAACCTTCGCTTCCAAACACTGCTGTGAGAAAGTGCCTCAAGTTCCTCTCTGC	838
QY	124	SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer	143
DB	839	TCTGAACACCTACTGTGACTTTATCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC	898
QY	144	PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln	163
DB	899	TTCTTGAGAGAAATATCGACCAATTAAGCTTTACATCAGTATGCACTCATACTCCCAA	958
QY	164	HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer	183
DB	959	CAATACTGTTCCCTATTCCTATTAACAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG	1018
QY	184	LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr	203
DB	1019	CTAGTGGCCAGCAGAGTTCGTGCAATTGGAAGTTTTTAATAAAACACACAGGTACACA	1078
QY	204	HisGlyHisGlySerGlnThrLeuTyrLeuAlaProGlyGlyAspAspTrpIleTyr	223
DB	1079	CACGGCAGTGGCTCAGAAAAGTTTTATATCTAGCTCCTGGAGGTTCTGACGATGGATCTAT	1138

Search completed: January 2, 2004, 00:16:40
Job time : 3811.88 secs

This Page Blank (uspto)